

OTDK dolgozat

2015

Regressziós játékok és alkalmazásaik

Regression games and applications

Kézirat lezárva: 2014. június 19.

Contents

1	Introduction	1
2	Cooperative games	3
2.1	Transferable utility games	3
2.2	Properties of the Shapley value	5
2.3	Regression games	9
2.3.1	The Shapley value as a measure of relative importance	12
3	Applications	15
3.1	Autoregressive games	15
3.1.1	The class of autoregressive games	16
3.1.2	The Shapley model selection of autoregressive processes	19
3.2	Least absolute deviation games	24
3.2.1	The class of least absolute deviation games	24
3.3	Binary regression games	28
3.3.1	The class of binary regression games	28
3.3.2	Goodness of fit	30
4	Summary	34
	Bibliography	36
	Appendix	I
A	Cooperative games	I
A.1	The <i>randlrg</i> MATLAB function	I
A.2	The <i>reallrg</i> MATLAB function	V
B	Applications	VI
B.1	Autoregressive simulation in Example 3.10	VI
B.2	Autoregressive correlation matrices in Example 3.10	VI

B.3	Autoregressive simulation in Example 3.12	VII
B.4	Autoregressive correlation matrices in Example 3.12	VII
B.5	LAD game payoffs in Example 3.22	IX
B.6	Regression game payoffs in Example 3.22	X
B.7	Logit regression game payoffs in Example 3.33	XI

List of Figures

2.1	The coalitions of a model with 3 predictors	10
3.1	The Shapley model selection of autoregressive processes	21
3.2	LAD optimization problem with infinite solutions	25
3.3	Shapley values and singleton coalition payoffs in Example 3.33	32

List of Tables

2.1	The marginal contributions of the players in Example 2.10	5
2.2	The marginal contributions of the players in Example 2.16	7
2.3	The marginal contributions of the players in Example 2.20	8
2.4	The marginal contributions of the players in Example 2.22	9
B.1	LAD game payoffs of the coalitions in Example 3.22	IX
B.2	Regression game payoffs of the coalitions in Example 3.22	X
B.3	Logit regression game payoffs of the coalitions in Example 3.33	XI

Chapter 1

Introduction

Decomposing goodness of fit measures and the relative importance of predictor variables are central problems in the theory of regression analysis. The relative importance of the predictor variables is quite obvious when the predictor variables are uncorrelated. Nevertheless, when the predictor variables are correlated finding the relative order of importance is a complex problem. There are quite straightforward relative importance measures that can be used when the predictor variables are correlated like partial goodness of fit values. However, the widely applied relative importance measures (Grömping, 2006, 2007; Lipovetsky and Conklin, 2005) have drawbacks that limit their area of utilization. To mention just few problems, some decompositions in use assign negative shares to variables while others assign positive shares to predictor variables that are uncorrelated with the predicted variable.

The main question of the paper is how we can decompose measures of goodness of fit in different types of regression models according to the Shapley value (Shapley, 1953), and how can we use the Shapley value as a relative importance measure. Our hypothesis, is that if certain conditions hold, the goodness of fit measures can be decomposed into relative importance shares with the Shapley value. We assume that the Shapley value as a relative importance measure has properties that make it useful for model selection.

We argue that the shares might serve as relative importance measures of discrete choice, least absolute deviation and autoregressive models. The fact that the Shapley value is a proper relative importance measure in linear regression is well known, and different approaches of this problem were done by Kruskal (1987); Feldman (2000); Lipovetsky and Conklin (2001) and Pintér (2006, 2007, 2011). Each of these works has a distinctive approach of the relative importance measuring with Shapley value. Some of these works does not use the axioms, definitions and methods of cooperative game theory, but unknowingly uses the Shapley value for goodness of fit decomposition. This approach is called averaging-over-orders in the econometric literature (Kruskal, 1987). An axiomatic

argument of the Shapley value used as relative importance measure is done before the introduction of the previously mentioned goodness of fit decompositions. Some of these axioms correspond to certain expected beneficial properties of a goodness of fit decomposition. We also argue that the issues, that are mentioned by Grömping (2007) as arguments against the use of Shapley value as a relative importance measure are closely related to the estimation of the random variables. Our approach with the known random variables guarantees that this problem does not appear.

Those paper that consider an axiomatic approach and use a cooperative game theory framework are only restricted to linear regression modeling. However, applications of the Shapley value for model selection and relative importance measuring in non-linear models were successfully done earlier by Lipovetsky and Conklin (2004); Huettner and Sunder (2012). Based on the relative importance measuring of predictor variables in the non-linear regression models we might also do relative importance orderings of the predictor variables. Such relative importance orders might be used for selection of the predictor variables with the greatest modeling potential. In addition in the paper we will introduce a stepwise model selection method of autoregressive processes, that could select the most important lagged predictors in a discrete time series.

The structure of the paper is the following: in Chapter 2 we introduce some axioms and a framework of cooperative game theory. We present the Shapley value, a solution concept of cooperative game theory and with examples we show the properties of the Shapley value. We introduce regression games, a class of transferable utility games. In Chapter 3 we introduce three new classes of regression games with examples: autoregressive games, least absolute deviation games and binary regression games. We show some properties of these games, and we apply the Shapley value for goodness of fit decomposition. In Chapter 4 we summarize our contributions.

Chapter 2

Cooperative games

First our goal is to introduce cooperative transferable utility games, and to demonstrate the properties of regression games. We have to suppose that, the reader is unfamiliar with game theory, and due to the interdisciplinary nature of the paper, this chapter introduces some cooperative game theory concepts.

2.1 Transferable utility games

We introduce transferable utility games, and their properties – on the basis of this game theory framework we can define new classes of regression games, and later in the next chapter with examples we can show the properties of these games.

Definition 2.1. *Let N be a finite set of the players, and $v : \mathcal{P}(N) \rightarrow \mathbb{R}$ be a function, such that $v(\emptyset) = 0$, where $\mathcal{P}(N)$ is the power set of N . Then we call v transferable utility (TU) cooperative game, henceforth game.*

The interpretation of the previous definition is straightforward, the subsets of N are the coalitions, and the value $v(S)$ is the payoff of coalition $S \in \mathcal{P}(N)$. The payoffs can be considered as money (Forgó et al., 2006), or anything else that can be divided, and the members of the coalitions can arbitrarily distribute the payoffs among each other.

Definition 2.2. *The game $v \in \mathcal{G}^N$ is monotone, if for all $A, B \in \mathcal{P}(N)$ it holds that, $A \subseteq B$, implies $v(A) \leq v(B)$.*

Example 2.3. Let $N = \{1, 2, 3\}$, and let v be the following monotone game:

$$\begin{aligned} v(\{1\}) &= 2 & v(\{2\}) &= 2 & v(\{3\}) &= 2 \\ v(\{1, 2\}) &= 5 & v(\{1, 3\}) &= 5 & v(\{2, 3\}) &= 5 \\ v(\{1, 2, 3\}) &= 9 \end{aligned}$$

Definition 2.4. The game $v \in \mathcal{G}^N$ is superadditive, if for all $A, B \in \mathcal{P}(N)$ such that $A \cap B = \emptyset$, $v(A \cup B) \geq v(A) + v(B)$.

The game from Example 2.3 is a superadditive game.

Definition 2.5. The game $v \in \mathcal{G}^N$ is subadditive, if for all $A, B \in \mathcal{P}(N)$ such that $A \cap B = \emptyset$, $v(A \cup B) \leq v(A) + v(B)$.

Example 2.6. Let $N = \{1, 2, 3\}$, and let v be the following subadditive game:

$$\begin{aligned} v(\{1\}) &= 2 & v(\{2\}) &= 2 & v(\{3\}) &= 2 \\ v(\{1, 2\}) &= 3 & v(\{1, 3\}) &= 3 & v(\{2, 3\}) &= 3 \\ v(\{1, 2, 3\}) &= 4 \end{aligned}$$

Definition 2.7. The game $v \in \mathcal{G}^N$ is additive, if for all $A, B \in \mathcal{P}(N)$ it holds that $A \cap B = \emptyset$, implies $v(A \cup B) = v(A) + v(B)$.

An additive game is superadditive and subadditive.

Definition 2.8. The game $v \in \mathcal{G}^N$ is essential (von Neumann and Morgenstern, 1944), if $v(N) > \sum_{i \in N} v(\{i\})$.

The game from Example 2.3 is an essential game.

Definition 2.9. For the game $v \in \mathcal{G}^N$, let $v'(S) = v(S \cup \{i\}) - v(S)$, that is, $v'(S)$ is player i 's marginal contribution to coalition S in game v , where $i \in N$, $S \in \mathcal{P}(N)$. Furthermore for any $i \in N$ let

$$D_{Sh}^i(S) = \begin{cases} \frac{|S|!(|N| - |S| - 1)!}{|N|!}, & \text{if } i \notin S \\ 0 & \text{otherwise} \end{cases}$$

a probability distribution on $\mathcal{P}(N)$. Then the Shapley value (Shapley, 1953) of player i in game v is the following:

$$Sh_i(v) = \sum_{S \in \mathcal{P}(N)} v'(S) D_{Sh}^i(S) \quad (2.1)$$

The Shapley value can be explained as follows, we list every permutation of the players; with the players set N we have $|N|!$ permutations. We calculate the marginal contributions of the players for each permutation. The arithmetic mean of these marginal contributions for each player is the Shapley value.

Example 2.10. Let be $N = \{1, 2, 3\}$, and let v be the following:

$$\begin{aligned} v(\{1\}) &= 5 & v(\{2\}) &= 12 & v(\{3\}) &= 15 \\ v(\{1, 2\}) &= 17 & v(\{1, 3\}) &= 18 & v(\{2, 3\}) &= 20 \\ v(\{1, 2, 3\}) &= 23 \end{aligned}$$

The permutations are the following:

$$\begin{aligned} \mathbf{1. arrival} & \quad 1 \quad 1 \quad 2 \quad 2 \quad 3 \quad 3 \\ \mathbf{2. arrival} & \quad 2 \quad 3 \quad 1 \quad 3 \quad 1 \quad 2 \\ \mathbf{3. arrival} & \quad 3 \quad 2 \quad 3 \quad 1 \quad 2 \quad 1 \end{aligned}$$

We can calculate the marginal contributions for each permutation:

	Marginal contributions					
	1.	2.	3.	4.	5.	6.
1	5	5	5	3	3	3
2	12	5	12	12	5	5
3	6	13	6	8	15	15

Table 2.1: The marginal contributions of the players in Example 2.10

The Shapley values of the players:

$$\begin{aligned} Sh_1(v) &= \frac{1}{6} \cdot (5 + 5 + 5 + 3 + 3 + 3) = 4 \\ Sh_2(v) &= \frac{1}{6} \cdot (12 + 5 + 12 + 12 + 5 + 5) = \frac{17}{2} \\ Sh_3(v) &= \frac{1}{6} \cdot (6 + 13 + 6 + 8 + 15 + 15) = \frac{63}{6} \end{aligned}$$

2.2 Properties of the Shapley value

First we introduce the concept of solutions:

Definition 2.11. Let $A \subseteq \mathcal{G}^N$, the function $\phi : A \rightarrow \mathbb{R}^N$ is a solution on A .

The previously introduced Shapley value is a solution, that has distinctive properties. The relevance of these properties is discussed with detailed examples, because the goodness of fit decompositions that will be introduced in the next sections inherit these properties. The profound understanding of them is essential, when we want to apprehend

and compare the properties of the relative importance measures. The Shapley value holds the following properties (among others):

1. Pareto optimal (PO) (Efficiency)
2. Equal treatment property (ETP)
3. Null player property (NP)
4. Strong monotonicity (SM)

According to Young (1985) properties 1, 2 and 4 axiomatize the Shapley value on $\mathcal{G}^N$. The Shapley value is the only solution, that holds properties 1, 2 and 4 on $\mathcal{G}^N$. Property 3 is a part of Shapley's original axiomatization (Shapley, 1953), it's introduction is not necessary, but later it will be relevant. Another nice property of the Shapley value is that, it can be calculated based on the payoffs. Some other solution concepts (Forgó et al., 2006), need linear programming or other modeling.

Definition 2.12. *The solution ϕ on $A \subseteq \mathcal{G}^N$, is Pareto optimal, if for each $v \in A$, it satisfies $\sum_{i \in N} \phi_i(v) = v(N)$.*

Example 2.10 shows this property:

$$Sh_1(v) + Sh_2(v) + Sh_3(v) = 4 + \frac{17}{2} + \frac{63}{6} = 23 = v(\{1, 2, 3\})$$

Interpretation of this property is simple, the Shapley value vector sums up to the value of the grand coalition.

Definition 2.13. *Let $v \in \mathcal{G}^N$, $i \in N$, and for each $S \subseteq N$: let $v'_i(S) = v(S \cup \{i\}) - v(S)$. Then v'_i is called Player i 's marginal contribution function in game v .*

This definition is needed, for the the introduction of equivalence sets, and the strong monotonicity property.

Definition 2.14. *Let $v \in \mathcal{G}^N$ be a game. Players $i, j \in N$ are equivalent in game v , $i \sim^v j$, if for all $S \subseteq N$ such that $i, j \notin S$: $v'_i(S) = v'_j(S)$. S is an equivalence set in v if for all $i, j \in S$: $i \sim^v j$.*

This definition is needed for the introduction of the Shapley value's ETP property.

Definition 2.15. *The solution ϕ on $A \subseteq \mathcal{G}^N$ satisfies the equal treatment property, if for all $v \in A$, $i, j \in N$ such that $i \sim^v j$ implies $\phi_i(v) = \phi_j(v)$.*

The ETP property of the Shapley value guarantees, that two players who always have the same marginal contributions to every coalition, not including them, are treated equally, that is, receive the same valuation. It should be noted, that the reversal is not true - two players with the same Shapley value not necessarily have the same contributions to every coalition.

Example 2.16. Consider the following game, let $N = \{1, 2, 3\}$, and v be the following:

$$\begin{aligned} v(\{1\}) &= 5 & v(\{2\}) &= 5 & v(\{3\}) &= 10 \\ v(\{1, 2\}) &= 15 & v(\{1, 3\}) &= 18 & v(\{2, 3\}) &= 18 \\ v(\{1, 2, 3\}) &= 30 \end{aligned}$$

We can calculate the contributions for every permutation:

Player	Marginal contributions					
	1.	2.	3.	4.	5.	6.
1	5	5	10	12	8	12
2	10	12	5	5	12	8
3	15	13	15	13	10	10

Table 2.2: The marginal contributions of the players in Example 2.16

The Shapley value of each player:

$$\begin{aligned} Sh_1(v) &= \frac{1}{6} \cdot (5 + 5 + 10 + 12 + 8 + 12) = \frac{26}{3} \\ Sh_2(v) &= \frac{1}{6} \cdot (10 + 12 + 5 + 5 + 12 + 8) = \frac{26}{3} \\ Sh_3(v) &= \frac{1}{6} \cdot (15 + 13 + 15 + 13 + 10 + 10) = \frac{38}{3} \end{aligned}$$

It stands that $v(\{1\}) = v(\{2\})$ and $v(\{1, 3\}) = v(\{2, 3\})$, this implies that Players 1 and 2 are symmetrical, so they have the same Shapley value.

Definition 2.17. Player i is a null player in game $v \in \mathcal{G}^N$, if $v(S \cup \{i\}) = v(S)$, for each $S \subseteq N$.

Example 2.18. Consider the following game, let $N = \{1, 2, 3\}$, and let v be the following cooperative TU game:

$$\begin{aligned} v(\{1\}) &= 0 & v(\{2\}) &= 5 & v(\{3\}) &= 10 \\ v(\{1, 2\}) &= 5 & v(\{1, 3\}) &= 10 & v(\{2, 3\}) &= 18 \\ v(\{1, 2, 3\}) &= 18 \end{aligned}$$

The contribution of player 1 is always 0 – player 1 is a null player.

Definition 2.19. The solution ϕ on $A \subseteq \mathcal{G}^N$ meets the null player property, if for all $v \in A, i \in N$ being a null player implies $\phi_i(v) = 0$.

Example 2.20. Example 2.18 shows the null player property. The marginal contributions of player 1 are always 0, so based on the payoffs we can calculate the marginal contributions for every order of arrival:

Player	Marginal contributions					
	1.	2.	3.	4.	5.	6.
1	0	0	0	0	0	0
2	5	8	5	5	8	8
3	13	10	13	13	10	10

Table 2.3: The marginal contributions of the players in Example 2.20

Based on the marginal contributions, we can calculate the Shapley values:

$$\begin{aligned}
 Sh_1(v) &= \frac{1}{6} \cdot (0 + 0 + 0 + 0 + 0 + 0) = 0 \\
 Sh_2(v) &= \frac{1}{6} \cdot (5 + 8 + 5 + 5 + 8 + 8) = \frac{13}{2} \\
 Sh_3(v) &= \frac{1}{6} \cdot (13 + 10 + 13 + 13 + 10 + 10) = \frac{23}{2}
 \end{aligned}$$

Player 1 always has zero marginal contribution, as a result of this he has Shapley value that is 0. This is one of the properties that is relevant later in the decomposition of goodness of fit values. In the reduced game containing only player 2 and 3 the players would receive the same Shapley values.

Definition 2.21. The solution ϕ on $A \subseteq \mathcal{G}^N$ is strongly monotone, if for all $v, w \in A, i \in N, v'_i \leq w'_i$, implies $\phi_i(v) \leq \phi_i(w)$.

Example 2.22. Consider the following game, let $N = \{1, 2, 3\}$, and w be the following game:

$$\begin{aligned}
 w(\{1\}) &= 7 & w(\{2\}) &= 5 & w(\{3\}) &= 10 \\
 w(\{1, 2\}) &= 17 & w(\{1, 3\}) &= 20 & w(\{2, 3\}) &= 18 \\
 w(\{1, 2, 3\}) &= 32
 \end{aligned}$$

We can calculate the marginal contributions for every order of arrival:

Player	Marginal contribution					
	1.	2.	3.	4.	5.	6.
1	7	7	12	14	10	14
2	10	12	5	5	12	8
3	15	13	15	13	10	10

Table 2.4: The marginal contributions of the players in Example 2.22

The Shapley values:

$$\begin{aligned}
Sh_1(w) &= \frac{1}{6} \cdot (7 + 7 + 12 + 14 + 10 + 14) = \frac{32}{3} \\
Sh_2(w) &= \frac{1}{6} \cdot (10 + 12 + 5 + 5 + 12 + 8) = \frac{26}{3} \\
Sh_3(w) &= \frac{1}{6} \cdot (15 + 13 + 15 + 13 + 10 + 10) = \frac{38}{3}
\end{aligned}$$

We can compare the Shapley values with Example 2.16. The only difference between the games given in Examples 2.16 and 2.22 is that Player 1's marginal contribution function is always greater in this game. This can be formalized like as follows: $v'_1 \leq w'_1$. This implies that $Sh_1(v) \leq Sh_1(w)$.

2.3 Regression games

In this section we introduce the notion of regression games, by regression we mean linear regression. Later we define properly the classes of regression games. Our goal is to measure the relative importance of the predictor variables of the model. The formalization and the axiomatization of the original problem is based on Pintér (2006), Pintér (2007) and Pintér (2011).

Let η be the dependent random variable, and $\xi_i, i = 1, \dots, n$ are the predictor variables. We suppose that the random variables are known, and we do not have to estimate them.

Definition 2.23. Let $N = \{\xi_1, \dots, \xi_n\}$ be the players set of the n predictor variable.

The coalitions, that are the subsets of N , resemble the groups of the random variables. Such groups of the variables can be demonstrated as follows:

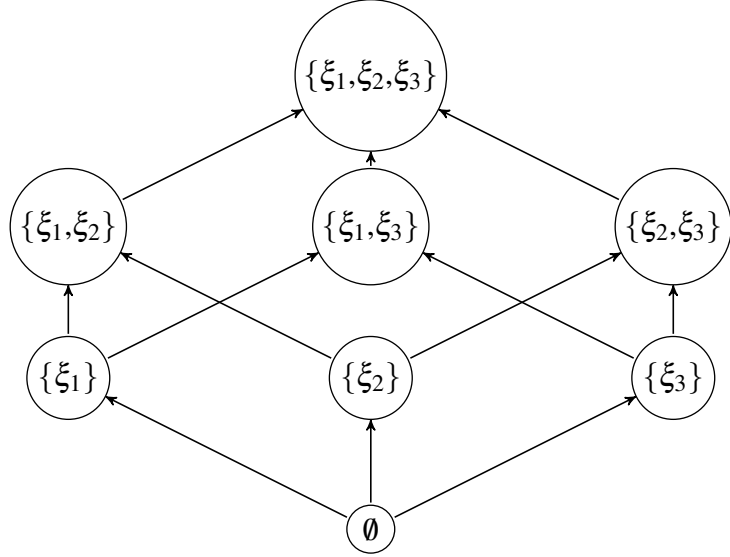


Figure 2.1: The coalitions of a model with 3 predictors

In the following we assume that N is fixed, which means that the full model includes n predictor variables, and the empty model includes no predictor variables. Let us consider the following optimization problem:

$$\begin{aligned} \text{var}(\eta) - \text{var} \left(\eta - \sum_{i \in S} \beta_i \cdot \xi_i \right) &\rightarrow \max \\ \text{s.t. } \beta_i &\in \mathbb{R}, \quad i \in S \end{aligned} \quad (2.2)$$

Definition 2.24. Let η be the predicted variable, and $\xi_1, \dots, \xi_n$ be the predictor variables. For any $S \in \mathcal{P}(N)$, let $v(S)$ be the solution of (2.2).

The payoffs of the coalitions are defined by the goodness of fit (GOF) in the model. The measure of GOF in this model is not the explained sum of squares (ESS). Explained sum of squares is equivalent with sum of squares due to regression (SSR). The goodness of fit is the difference of the variances, that is maximized. When we divide the ESS of a linear regression model with the total sum of squares (TSS), we achieve the multiple coefficient of determination (R^2). The problem is that, TSS and ESS is only meaningful, when we have to estimate the random variables. However, R^2 of the linear regression models can be calculated from the correlation matrix or from the covariance matrix. The use of different GOF measures of a linear regression model is a matter of debate in the literature (Pintér, 2007; Lipovetsky and Conklin, 2001), but it can be easily seen that the normalization is not a problem, when we want to compare the payoffs (N and η are fixed).

Corollary 2.25. Function v in Definition 2.24 is a TU game.

We take the predictor variables as the players, and the sets of variables in the model are the coalitions. The GOF values give the payoffs of the coalitions of N .

Definition 2.26. We call the games defined by Definitions 2.23 and 2.24 linear regression games. The class of linear regression games is denoted by $\mathcal{G}_{LR}^N$.

Each type of regression models e.g. logit, least absolute deviation, linear regression is based on optimization problems. The difference among these classes of regression games is essential because the definition and the set up of the optimization problem is different in each case, therefore the definition of GOF is also different.

Proposition 2.27. The class of $\mathcal{G}_{LR}^N$ is a subset of the class of monotone games.

An additional predictor variable in the model never decreases the explained variance (Wooldridge, 2012), it follows, that the payoffs never decrease when a new player joins to the coalition. This implies, that the games in $\mathcal{G}_{LR}^N$ are monotone.

Example 2.28. Consider the following regression model, where we observe the properties of some cars – the database is from Kane (2002):

$$\eta = \beta_0 + \sum_{i=1}^3 \beta_i \cdot \xi_i,$$

where η is the manufacturers suggested retail price in USD, and the predictor random variables are the following:

ξ_1 :	Vehicle weight in pound	(Weight)
ξ_2 :	Power of engine in horsepower	(Power)
ξ_3 :	1 if the vehicle is compact	(Compact)

We define ε as the error random variable like the following:

$$\varepsilon \equiv \eta - \beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i$$

Let $N = \{\xi_1, \xi_2, \xi_3\}$ be the set of players, and let the R^2 value be the measure of goodness of fit. Without the calculation of parameters, only based on the correlation matrix we can calculate the R^2 values of the coalitions.

The following **C** correlation matrix contains the correlations between ξ_1, ξ_2, ξ_3 and η :

$$\mathbf{C} = \begin{matrix} & \eta & \xi_1 & \xi_2 & \xi_3 \\ \begin{matrix} \eta \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{matrix} & \begin{pmatrix} 1 & 0.61 & 0.84 & -0.48 \\ 0.61 & 1 & 0.43 & -0.59 \\ 0.84 & 0.43 & 1 & -0.55 \\ -0.48 & -0.59 & -0.55 & 1 \end{pmatrix} \end{matrix}$$

The payoffs of the different coalitions¹ are the following:

$$\begin{aligned} v(\{\xi_1\}) &\approx 0.3721 & v(\{\xi_2\}) &\approx 0.7056 & v(\{\xi_3\}) &\approx 0.2304 \\ v(\{\xi_1, \xi_2\}) &\approx 0.7815 & v(\{\xi_1, \xi_3\}) &\approx 0.3942 & v(\{\xi_2, \xi_3\}) &\approx 0.7061 \\ v(\{\xi_1, \xi_2, \xi_3\}) &\approx 0.7964 \end{aligned}$$

This example also shows that, games in $\mathcal{G}_{LR}^N$ are not necessarily superadditive or sub-additive. The Shapley values of the random variables, in this specific model are the following:

$$Sh_1(v) \approx 0.1940 \quad Sh_2(v) \approx 0.5168 \quad Sh_3(v) \approx 0.0856$$

The Shapley values show that the power of engine is the most important predictor variable, followed by the vehicles weight in pound, and the compact car design is the least important. Based on the fact, that the Shapley values are not equivalent with the squared correlation values between the predictors and the predicted variable, we can assume that collinearity among the predictors is present. If the predictors would be uncorrelated with each other, they would receive their squared correlation with the predicted variable as a Shapley value (Pintér, 2007).

2.3.1 The Shapley value as a measure of relative importance

Relative importance measures decompose a measure of GOF e.g. R^2 , log-likelihood, sum of absolute deviation into so called GOF shares, and assign a share for each predictor variable, according to the relevance of the variable. Significantly different relative importance measures are used in the literature, but each of them has certain weaknesses. Some problems with these relative importance measures, are as follows:

1. Products of standardized coefficients with marginal contributions and the method of incremental net effects (Lipovetsky and Conklin, 2005) can allocate negative shares to the explanatory variables. This is problematic because a negative relative importance implies that the predictor variable decreases the GOF. An additional predictor variable never decreases the GOF of a linear regression, least absolute deviation or logit model.
2. The sum of partial R^2 values (Hajdu, 2013a) and the sum of some other relative importance measures (Kvalseth, 1985) are not always equivalent with the chosen GOF that is being decomposed. This violates the property *Non negativity*.

¹The codes of the MATLAB programs that calculated the payoff vector and the Shapley value vector in the example are included in Section A.1 and A.2 of the Appendix.

3. The R^2 partitioned by proportional marginal variance decomposition (Grömping, 2007), assigns positive shares to predictors that are correlated with the other predictor variables, but uncorrelated with the predicted variable.

According to Grömping (2007) a proper relative importance measure should have the following properties:

1. *Proper decomposition*: the model variance is to be decomposed into shares, that is, the sum of the shares has to be the model variance. The 1. problem mentioned violates this.
2. *Non negativity*: All shares have to be non-negative. The 2. problem mentioned violates this.
3. *Exclusion*: The share allocated to a regressor ξ_i with $\beta_i = 0$ should be 0.
4. *Inclusion*: The share allocated to a regressor ξ_i with $\beta_i \neq 0$ should receive a non zero share.

Proper decomposition, non negativity and inclusion are always fulfilled by the Shapley value (Huettner and Sunder, 2012). The first desired property – *proper decomposition* – is equivalent with the PO property of Shapley value applied on $\mathcal{G}_{LR}^N$. Because of this property of the Shapley value, every GOF measure can be *decomposed properly* with it, not just the games in $\mathcal{G}_{LR}^N$.

The second property – *non negativity* – is always fulfilled, when we define a GOF measure that is always positive, because in such cases the game is monotone, the elements of the marginal contribution will be also positive, so the Shapley values are positive. If needed, a monotone transformation of the GOF measure can be applied, and negative GOF measure values can be transformed into positives. (In the case of log-likelihood and $-2\log$ -likelihood, similar transformation is applied.)

The third property – *exclusion* – formulates, that the ξ_i explanatory variable with $\beta_i = 0$ receives a 0 share. There are many cases when a β_i that is 0 does not indicate a predictor that is uncorrelated with the predicted variable. For example in our not empirical approach, it is possible that a predictor variable ξ_i which is correlated with the predicted variable, and might be a linear combination of other predictors, could be in the model. In this case it might receive a β_i that is 0, but a Shapley value that is different from 0. This problem that is connected to parameter estimation would not rise in an empirical model, where the parameters are estimated with ordinary least squares (perfect

multicollinearity). The reason is that the correlation matrix would have no inverse, and we could not obtain the parameters. Therefore, in such situation the fact that the Shapley value is 0, not necessarily means that ξ_i is uncorrelated with predicted variable. However, the NP property of the Shapley value ensures, that a predictor variable uncorrelated with the predicted variable, has a 0 Shapley value. This phenomenon highlights that β parameters have less to do with relative importance.

The property – *inclusion* – formulates, that the ξ_i explanatory variable with $\beta_i \neq 0$ has marginal contributions to the models, and ξ_i is not a null player in the game. A player with a contribution different from 0 will receive a positive Shapley value.

Chapter 3

Applications

The definition of regression games, restricts the idea of regression to linear regression. However, the concept of regression games can be broadened, and the key of this broader approach is the proper definition of the random variables used in the models and the set up of the optimization problem.

Definition 3.1. *In this chapter we assume that the random variables are $\xi : \mathbb{N} \rightarrow \mathbb{R}$ functions.*

The introduction of such random variables is needed for the proper axiomatization of autoregressive games.

Definition 3.2. *We define the $k \in \mathbb{N}$ lag of the random variable ξ at the point x like as follows:*

$$\xi_{-k}(x) = \xi(x + k)$$

This approach of random variable lags is used in the following 3 sections.

3.1 Autoregressive games

One of the main assumptions of linear regression modeling is that, the error variable is uncorrelated with the predictor variables. We also suppose, that the errors for every k lag are uncorrelated. If this assumption fails we might face the following problems (Hajdu, 2013b) :

1. Biased GOF measures
2. Inconsistent errors
3. Inefficient parameter estimations

The problems 2 and 3 are only relevant when the random variables are estimated, but in this model variables are known. Biased GOF measures can directly affect the model selection method, because some of the model selection methods are based on the GOF values (e.g. information criteria). When this assumption fails the model suffers from residual autocorrelation. The cause of residual autocorrelation can be the following:

1. Misspecification of regression model
2. Absence of relevant variables
3. Low number of observations
4. Distributed lag process

The low number of observations only causes residual autocorrelation when the variables are estimated. Because we do not estimate the variables, only the cases described by point 1, 2 and 4 can be the cause of residual autocorrelation in this approach.

3.1.1 The class of autoregressive games

Consider the following linear regression model:

$$\eta = \beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i \quad (3.1)$$

The random variable η is the dependent random variable, $\xi_i, i = 1, \dots, n$ are the predictor variables and ε is the error random variable that is defined by the following:

$$\varepsilon \equiv \eta - \beta_0 - \sum_{i=1}^n \beta_i \cdot \xi_i$$

One of the main assumptions of ordinary least squares regression modeling is that the residuals are uncorrelated, what can be formalized as follows:

$$E(\varepsilon, \varepsilon_{-k}) = 0, \quad k \in \mathbb{N}$$

Where ε is the error variables realization, and k is the lag between the realizations.

Definition 3.3. *The deviation of the error variable, for a k lag can be achieved as follows:*

$$\sigma_{\varepsilon-k} = \sqrt{COV(\varepsilon, \varepsilon_{-k})}$$

Definition 3.4. *We can calculate the residual-autocorrelation coefficient ϕ_k within ε for a k lag like as follows:*

$$\phi_k = \frac{COV(\varepsilon, \varepsilon_{-k})}{\sigma_{\varepsilon} \cdot \sigma_{\varepsilon-k}}, \quad k \in \mathbb{N}$$

Definition 3.5. *The model defined by (3.1) suffers from residual autocorrelation if the following stands:*

$$|\phi_k| > 0, \quad k \in \mathbb{N}$$

Instead of 0, an arbitrary δ cut value can be used:

$$|\phi_k| > \delta, \quad k \in \mathbb{N}, \quad \delta \in (0, 1)$$

First we assume that the process is not a distributed lag process and the model is well specified, e.g. the function is neither logarithmic, nor exponential and so on. We also assume that every exogenous variable is in the model. Including lagged dependent variables can solve the residual autocorrelation problem, but identifying the order of the lags is not simple (Hannan and Quinn, 1979). The application of the Shapley value for lag selection can be supported, because the previously used general regression game model measures the same distances, and the solution of the optimization problem can be achieved by the same way. This ensures that it can be modified, without a new axiomatization.

Let η be the dependent random variable, $\xi_i, i = 1, \dots, n$ be the independent predictor variables, and $\eta_{-j}, j = 1, \dots, k$ be the lagged dependent variables.

Definition 3.6. *Let $N = \{\xi_1, \dots, \xi_n, \eta_{-1}, \dots, \eta_{-k}\}$ be the players sets of the $n + k$ predictors.*

In the following we assume that N is fixed, which means that the full model includes $n + k$ predictor variables, and the empty model includes no predictor variables. Let consider the following optimization problem:

$$\begin{aligned} \text{var}(\eta) - \text{var} \left(\eta - \sum_{i \in S} \beta_i \cdot \xi_i - \sum_{j \in S} \phi_j \cdot \eta_{-j} \right) &\rightarrow \max \\ \text{s.t. } \beta_i, \phi_j &\in \mathbb{R}, \quad i, j \in S \end{aligned} \quad (3.2)$$

Definition 3.7. *Let η be the predicted variable, and $\xi_1, \dots, \xi_n, \eta_{-1}, \dots, \eta_{-k}$ are the predictor variables. For any $S \in \mathcal{P}(N)$, let $v(S)$ be the solution of (3.2).*

The payoffs of the coalitions are defined again by the GOF value in the model. Because the lagged predictor variables, we do not have to use different space measuring, because of this the R^2 can be used as measure of GOF again.

Corollary 3.8. *Function v in Definition 3.7 is a game.*

We take the predictor variables and lagged variables as the players, and the sets of variables in the model as the coalitions. The GOF values give the payoffs of the coalitions of N .

Definition 3.9. We call the games defined by Definitions 3.6 and 3.7 autoregressive games. The class of autoregressive games is denoted by $\mathcal{G}_{AR}^N$.

We only define autoregressive games here, but every discrete moving average process can be rewritten into a discrete autoregressive process (Mikusheva, 2007), if certain conditions hold.

Example 3.10. Consider the following simple regression model - where the model can describe the value of a process over time:

$$\eta = \beta_0,$$

where η is the value of the process and we define ε as the random error variable as follows:

$$\varepsilon \equiv \eta - \beta_0$$

We assume that the error variable is autocorrelated. In this case, because of the empty model, which contains no predictors except the slope, it means that the process itself is autocorrelated. In the following we suppose that the process is autocorrelated for a high order of k , the exact order and the structure of the autocorrelation is unknown, we only know the correlation matrix of the predicted variable and the lags for an optional k order. We choose the arbitrary $k = 5$ lag order, so our augmented model is the following:

$$\eta = \beta_0 + \sum_{j=1}^5 \phi_j \cdot \eta_{-j}$$

We also assume that, the process is stationary, and there is no unit root in the model. This assumption is essential (Dickey and Fuller, 1979), and it can be formalized as follows:

$$\mu \equiv E(\eta) \tag{3.3}$$

$$\gamma_{-k} \equiv COV(\eta, \eta_{-k}), \quad k \in \mathbb{N} \tag{3.4}$$

Equation (3.3) formulates, that the process has a constant μ expected value, that does not change. This means that the expected value of the process is independent from time. Equation (3.4) formulates, that for a k lag the covariance of the process is a constant γ_{-k} value. This means that, the covariance only depends on the order of the lag.

Let $N = \{\eta_{-1}, \eta_{-2}, \eta_{-3}, \eta_{-4}, \eta_{-5}\}$ be the set of players, and the R^2 value be the measure of goodness of fit chosen as a payoff. Without the calculation of the parameters, only based on the correlation matrix we can compute the R^2 values of the models. The code

snippet used for simulation is attached in Section B.1 of the Appendix. The following $\mathbf{C}$ correlation matrix contains the correlations between $\eta_{-1}, \eta_{-2}, \eta_{-3}, \eta_{-4}, \eta_{-5}$ and η :

$$\mathbf{C} = \begin{matrix} & \eta & \eta_{-1} & \eta_{-2} & \eta_{-3} & \eta_{-4} & \eta_{-5} \\ \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \\ \eta_{-5} \end{matrix} & \begin{pmatrix} 1 & 0.74 & 0.61 & 0.6 & 0.7 & 0.64 \\ 0.74 & 1 & 0.74 & 0.61 & 0.6 & 0.7 \\ 0.61 & 0.74 & 1 & 0.74 & 0.61 & 0.6 \\ 0.6 & 0.61 & 0.74 & 1 & 0.74 & 0.61 \\ 0.7 & 0.6 & 0.61 & 0.74 & 1 & 0.74 \\ 0.64 & 0.7 & 0.6 & 0.61 & 0.74 & 1 \end{pmatrix} \end{matrix}$$

Based on the $\mathbf{C}$ matrix we can calculate the payoffs of the coalitions. The reduced correlation matrices are included in Section B.2 of the Appendix. From the payoff vector we can compute the Shapley values:

$$\begin{aligned} Sh_1(v) &\approx 0.2100 & Sh_2(v) &\approx 0.0893 & Sh_3(v) &\approx 0.0836 \\ Sh_4(v) &\approx 0.1635 & Sh_5(v) &\approx 0.1038 \end{aligned}$$

It can be assumed that from the predictor's set η_{-1} and η_{-4} are the most important predictors. The stepwise model selection method, that is introduced later will show that truly η_{-1} and η_{-4} are the most important predictors. However, the relevance of the other predictors and their contribution to the GOF value should not be considered low, seemingly $\eta_{-2}, \eta_{-3}, \eta_{-5}$ account for nearly 30% of the variance of η .

3.1.2 The Shapley model selection of autoregressive processes

On the grounds of the previously introduced class of games – $\mathcal{G}_{AR}^N$ – and based on the properties of the Shapley value and discrete stationary autoregressive processes we define a model selection algorithm of autoregressive time series.

The introduced method is a stepwise model selection method, which uses a backward elimination process. This means that the starting point of the model selection is the model which includes every predictor variable. Because the random variables are known, the method is not based on estimation, so it does not depend on significance levels. However, arbitrary cut values can be chosen for the model selection.

1. It is supposed that the time series process can be described, by a model that is augmented only with autoregressive lags of the dependent variable. This approach combined with stationary transformations allows us to select the lags of autoregressive processes. Our model can be written as follows

$$\eta = \beta_0 + \sum_{j=1}^k \phi_j \cdot \eta_{-j}, \quad j \in \mathbb{N}$$

2. We choose a k order of lags, and we obtain the correlation matrix $\mathbf{C}$ of the variables $\{\eta, \eta_{-1}, \dots, \eta_{-k}\}$.
3. In the model, $N = \{\eta_{-1}, \dots, \eta_{-k}\}$ is the player set of the k predictors. We define the payoffs as the GOF value, and we calculate the payoffs.
4. We calculate the Shapley values.
5. We order the variables in descending order, based on the relative importance of the variables.
6. We eliminate the least important predictor variable from the model and reduce the original correlation matrix, with this we obtain $\tilde{\mathbf{C}}$.
7. In the new model, with the reduced variable set is the player set of the $k - 1$ predictors. We define the payoffs as the GOF value, and we calculate the payoffs.
8. We compute the Shapley value for each player present in the reduced model, based on the $\tilde{\mathbf{C}}$ matrix.
9. If the payoff of the grand coalition remained the same, the elimination of the variable was well-founded. If the payoff of the grand coalition dropped, the elimination was not valid. This means that the chosen δ cut value is 0.
10. If the elimination was valid, we jump to point 3. Otherwise we remit the eliminated predictor and choose the next least important variable in the relative importance order for elimination.
11. The algorithm continues until no other predictor can be eliminated from the model.
12. When the algorithm terminated, we identified the $AR(k)$ model that describes best the process.

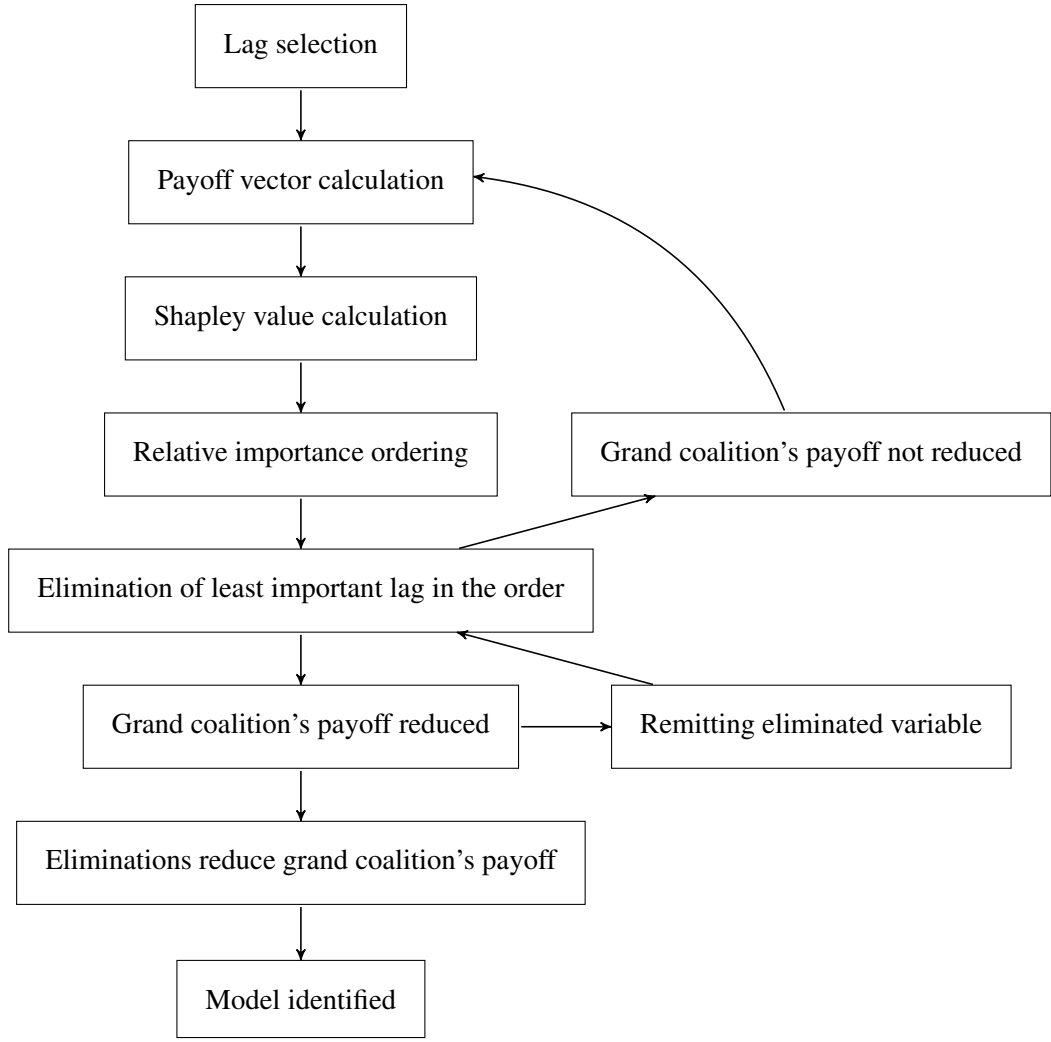


Figure 3.1: The Shapley model selection of autoregressive processes

Example 3.11. Let us consider again Example 3.10, and see what our algorithm gives. We reexamine the Shapley values, so here they are:

$$\begin{aligned}
 Sh_1(v) &\approx 0.2100 & Sh_2(v) &\approx 0.0893 & Sh_3(v) &\approx 0.0836 \\
 Sh_4(v) &\approx 0.1635 & Sh_5(v) &\approx 0.1038
 \end{aligned}$$

First we eliminate η_{-3} , because according to the Shapley values this has the lowest relative importance. Then based on the reduced correlation matrix – the reduced matrices are included in Section B.2 of the Appendix – the recalculated Shapley values are the following:

$$Sh_1(v) \approx 0.2296 \quad Sh_2(v) \approx 0.1081 \quad Sh_4(v) \approx 0.1917 \quad Sh_5(v) \approx 0.1207$$

The elimination of η_{-3} is confirmed by the fact that the sum of the Shapley value vector remained the same. Because the sum remained the same, the payoffs of the grand

coalition remained the same, so the GOF value was not reduced by the elimination. Now we eliminate η_{-2} , because it has the lowest share. The Shapley values are the following:

$$Sh_1(v) \approx 0.2795 \quad Sh_4(v) \approx 0.2237 \quad Sh_5(v) \approx 0.1469$$

Our last step of the selection method, is the elimination of η_{-5} , and based on the final correlation matrix the Shapley values are the following:

$$Sh_1(v) \approx 0.3538 \quad Sh_4(v) \approx 0.2962$$

The elimination of η_{-4} would reduce the payoff of the grand coalition, so our model is final and stable, the relative importance of η_{-1} and η_{-4} is proved. The original $\mathbf{C}$ correlation matrix belonged to a simulated autoregressive process with the following structure:

$$\eta = \beta_0 + \phi_1 \cdot \eta_{-1} + \phi_4 \cdot \eta_{-4} + \varepsilon, \quad \varepsilon \approx WN(\mu, \sigma)$$

The $WN(\mu, \sigma)$ denotes white noise with μ expected value and σ deviation. The parameters of the simulation were the following:

$$\beta_0 = 3 \quad \phi_1 = 0.5 \quad \phi_4 = 0.4 \quad \mu = 0 \quad \sigma = 4$$

Example 3.12. Let us consider the following problem, where the autoregressive problem's structure is as in Example 3.10. There we have the following correlation matrix for $k = 7$ lags:

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \\ \eta_{-5} \\ \eta_{-6} \\ \eta_{-7} \end{matrix} \begin{pmatrix} 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 & 0.81 & 0.51 \\ 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 & 0.81 \\ 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 \\ 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 \\ 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 \\ 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 \\ 0.81 & 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 \\ 0.51 & 0.81 & 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 \end{pmatrix}$$

Based on matrix $\mathbf{C}$ we can calculate the payoffs of the coalitions. From the payoff vector we can compute the Shapley values:

$$\begin{aligned} Sh_1(v) &\approx 0.0406 & Sh_2(v) &\approx 0.2520 & Sh_3(v) &\approx 0.0409 \\ Sh_4(v) &\approx 0.2310 & Sh_5(v) &\approx 0.0394 & Sh_6(v) &\approx 0.1728 \\ Sh_7(v) &\approx 0.0393 \end{aligned}$$

First we eliminate η_{-7} , because according to the Shapley values, this has the lowest relative importance. After this, based on the reduced correlation matrix – the reduced

matrices are included in Section B.4 of the Appendix – the recalculated Shapley values are the following:

$$\begin{aligned} Sh_1(v) &\approx 0.0466 & Sh_2(v) &\approx 0.2589 & Sh_3(v) &\approx 0.0473 \\ Sh_4(v) &\approx 0.2380 & Sh_5(v) &\approx 0.0457 & Sh_6(v) &\approx 0.1795 \end{aligned}$$

The elimination of η_{-7} is confirmed by the fact that the sum of the Shapley value vector did not change. Because the sum remained the same, the payoff of the grand coalition remained the same, the GOF value was not reduced by the elimination. Now we eliminate η_{-5} , because it has the lowest share. The recomputed Shapley values are the following:

$$\begin{aligned} Sh_1(v) &\approx 0.0554 & Sh_2(v) &\approx 0.2685 & Sh_3(v) &\approx 0.0563 \\ Sh_4(v) &\approx 0.2473 & Sh_6(v) &\approx 0.1885 \end{aligned}$$

The elimination of η_{-5} was right, because the payoff of the grand coalition remained did not change. Now we eliminate the predictor variable η_{-1} , the Shapley values are the following:

$$Sh_2(v) \approx 0.2826 \quad Sh_3(v) \approx 0.0703 \quad Sh_4(v) \approx 0.2614 \quad Sh_6(v) \approx 0.2016$$

The elimination of η_{-1} was justified, now we eliminate η_{-3} with this we obtain the following Shapley values:

$$Sh_2(v) \approx 0.3062 \quad Sh_4(v) \approx 0.2844 \quad Sh_6(v) \approx 0.2232$$

This elimination reduced the payoff of the grand coalition, we remit η_{-3} and eliminate η_{-6} with this we achieve the following Shapley values:

$$Sh_2(v) \approx 0.3728 \quad Sh_3(v) \approx 0.0918 \quad Sh_4(v) \approx 0.3512$$

Further elimination would reduces the payoff of the grand coalition, so the important lags are η_{-2} , η_{-3} and η_{-4} . Correlation matrix $\mathbf{C}$ belongs to a simulated autoregressive process with the following structure:

$$\eta = \beta_0 + \phi_2 \cdot \eta_{-2} + \phi_3 \cdot \eta_{-3} + \phi_4 \cdot \eta_{-4} + \varepsilon, \quad \varepsilon \approx WN(\mu, \sigma)$$

The parameters of the simulation were the following:

$$\beta_0 = 5 \quad \phi_2 = 0.5 \quad \phi_3 = 0.05 \quad \phi_4 = 0.4 \quad \mu = 0 \quad \sigma = 2$$

The MATLAB code of the simulation is included in Section B.3 of the Appendix.

Remark 3.13. *It worth emphasizing that statistical software packages Gretl and SPSS do not omit η_{-6} .*

3.2 Least absolute deviation games

The general linear regression models, that minimize the unexplained variance (or maximize the explained) and were introduced in Sections 2.3 and 3.1 suffer from a significant weakness. If the random variables are estimated, the models are not robust, it means the variance is sensitive to the bias of the outliers, and these distortions effect the parameter estimation. These bias can falsely flag the presence of heteroscedasticity (non constant standard deviation of the dependent variable) or imply wrong model specification falsely. To control this problem, that is closely connected to the empirical approach, another regression model can be used on data with outliers - the least absolute deviation regression (Chen et al., 2008), henceforth LAD regression. We introduce a general LAD model, that is specified with the arithmetic mean and not with quantiles (Hajdu, 2013a).

3.2.1 The class of least absolute deviation games

In this paper this is the first class of games that uses a different space measure in the set up of the optimization problem.

Definition 3.14. *We define a distance of the random variables ξ and η as:*

$$d(\xi, \eta) = E(|\xi - \eta|)$$

As a reminder, the distance holds the following: non-negativity, identity of indiscernibles, symmetry, and the triangle inequality. This definition of distance is needed for the set up of the optimization problem, that has to be solved to achieve the parameters. This distance of the variables is the L_1 norm or the so-called Taxicab-Manhattan norm (Krause, 1986).

Definition 3.15. *Let $N = \{\xi_1, \dots, \xi_n\}$ the player set of the n predictors.*

In the following we assume that N is fixed, which means that the full model includes n predictor variables, and the empty model includes no predictor variables. Lets consider the following optimization problem:

$$\begin{aligned} E(|\eta|) - E\left(\left|\eta - \sum_{i \in S} \beta_i \cdot \xi_i\right|\right) &\rightarrow \max \\ \text{s.t. } \beta_i &\in \mathbb{R}, \quad i \in S \end{aligned} \tag{3.5}$$

Definition 3.16. *Let η be the predicted variable, and $\xi_1, \dots, \xi_n$ be the predictor variables. For any $S \in \mathcal{P}(N)$, let $v(S)$ be the solution of (3.5).*

Corollary 3.17. *Function v in Definition 3.16 is a game.*

We take the predictor variables as the players, and the sets of variables in the model as coalitions. The GOF values give the payoffs of the coalitions of N .

Definition 3.18. We call the games defined by Definitions 3.15 and 3.16 least absolute deviation games. The class of least absolute deviation games is denoted by $\mathcal{G}_{LAD}^N$.

This class of games is generally different from $\mathcal{G}_{LR}^N$ and $\mathcal{G}_{AR}^N$ because of the applied distance are different. It follows that cardinal comparison of the solutions achieved is not meaningful for the same set of players, because the applied measures of distance are different. However, an ordinal relative importance comparison of the solutions can be done with the solution vectors.

Remark 3.19. We suppose that the optimization problem (3.5) has a unique solution. The idea behind this, is that the optimization problem (3.5) might have many solutions (Bloomfield and Steiger, 1980).

Example 3.20. Consider set $D \subseteq \mathbb{R}^2$, and the cardinality of the set is n . The coordinates of the points are denoted by x_i, y_i and we want to find the line segment $y(x)$ over the interval $x \in [-10, 10]$, that minimizes $\sum_{i=1}^n |y(x) - y_i|$.

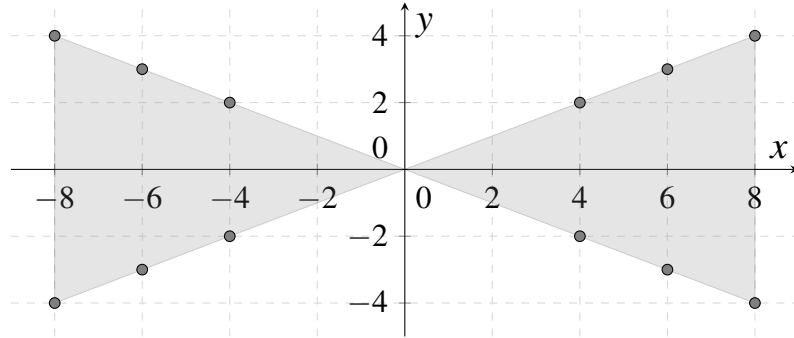


Figure 3.2: LAD optimization problem with infinite solutions

A randomly chosen set of points is denoted by the dots on Figure 3.2. The area shaded with gray shows the set of line segments, that satisfy the minimization problem. This example shows that occasionally the optimization problem that is formalized might have many solutions. However, this is not a problem because the models achieve the same optimal value. This also shows that the not standardized β values have nothing to do with relative importance.

Definition 3.21. We can define the coefficient of determination for LAD models as follows:

$$R_{LAD} \equiv \frac{E(|\eta|) - E\left(\left|\eta - \sum_{i \in S} \beta_i \cdot \xi_i\right|\right)}{E(|\eta|)} \quad (3.6)$$

This measure of GOF is similar to the coefficient of determination, that is used as measure of GOF in linear regression games. However, it cannot be calculated based on the correlation matrix of the predicted variable and the predictors (McKean and Sievers, 1987). Other measures of GOF that are widely used in LAD regression analysis are based on sample size, and the number of parameters estimated. Because we previously assume that the random variables are known, we do not consider this kind of GOF. The application of (3.6) as measure of GOF can be justified with the following reasons according to Kvalseth (1985):

1. The applied measure of GOF must be directly linked to the fitting criterion - the optimization problem itself. This is satisfied, because the numerator of (3.6) is (3.5).
2. The used measure of GOF must be invariant to the scale changes of η and the predictors. Because the measure of GOF defined by (3.6) is a fraction, this is also satisfied.
3. The measure of GOF must satisfy the following requirement: the value of the GOF has to be in a strict, previously defined interval. This measure of GOF satisfies, that $0 \leq \text{GOF} \leq 1$, with 0 meaning lack of fit (empty model), and 1 meaning the perfect fit.
4. The measure of GOF must never decrease when predictors are added to the model.
5. The measure of GOF should be robust. Outliers and the requirement of robustness is an empirical problem, our approach with the known random variables eliminates this problem.

Example 3.22. Consider the following LAD regression model, where we observe the properties of some cars – the database that is used can be found in Kane (2002):

$$\eta = \beta_0 + \sum_{i=1}^5 \beta_i \cdot \xi_i,$$

where η is the manufacturers suggested retail price in USD, and the predictor variables are the following:

ξ_1 :	Fuel usage efficiency in city, miles per gallon	(City)
ξ_2 :	Vehicle length in inch	(Length)
ξ_3 :	Vehicle height in inch	(Height)
ξ_4 :	Vehicle weight in pound	(Weight)
ξ_5 :	Power of engine in horsepower	(Power)

We define ε as the error variable:

$$\varepsilon \equiv \eta - \beta_0 + \sum_{i=1}^5 \beta_i \cdot \xi_i$$

Let $N = \{\xi_1, \xi_2, \xi_3, \xi_4, \xi_5\}$ be the set of players, and let the R_{LAD} values of the models be the payoffs. The detailed table containing the objective value functions, and the payoffs of the coalitions is included in the Appendix. Based on the payoffs, the Shapley values are as follows:

$$\begin{aligned} Sh_1(v) &\approx 0.0823 & Sh_2(v) &\approx 0.0403 & Sh_3(v) &\approx 0.0453 \\ Sh_4(v) &\approx 0.1433 & Sh_5(v) &\approx 0.254 \end{aligned}$$

Based on the Shapley values, we see that the engine power is the most important predictor. The next (in relative importance) is the vehicle weight, followed by the vehicle fuel usage in city environment. The variables, that describe the physical dimensions of the cars are the least important, they account for nearly 8% of the absolute deviation together. It is worth mentioning that the payoff share of the variable Height shows, that in a LAD regression game the Shapley value achieved by a player can be greater, than the player's individual payoff (for payoffs see Section B.5 of the Appendix). With the known random variables we have the following **C** correlation matrix, that includes the correlations between the variables $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5$ and η :

$$\mathbf{C} = \begin{matrix} & \begin{matrix} \eta \\ \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{matrix} & \begin{pmatrix} 1 & -0.60 & 0.33 & 0.08 & 0.60 & 0.84 \\ -0.60 & 1 & -0.47 & -0.45 & -0.80 & -0.69 \\ 0.33 & -0.47 & 1 & 0.21 & 0.63 & 0.45 \\ 0.08 & -0.45 & 0.21 & 1 & 0.69 & 0.08 \\ 0.60 & -0.80 & 0.63 & 0.69 & 1 & 0.61 \\ 0.84 & -0.69 & 0.45 & 0.08 & 0.61 & 1 \end{pmatrix} \end{matrix}$$

The correlation matrix above shows us that the predictor variable Height is lightly correlated with the most important predictor – Power. This relationship enlarges the importance of Height, and this is why it has a relative importance nearly as high as Length. We can compare the Shapley values of this model, with the ones that are achieved with the maximization of the explained variance. The set of predictors stays the same, and the predicted variable is also the same. Without estimating the random variables, based on the correlation matrix we can obtain the payoffs, (for payoffs see Section B.6 of the Appendix) and the Shapley values:

$$\begin{aligned} Sh_1(v) &\approx 0.1211 & Sh_2(v) &\approx 0.0483 & Sh_3(v) &\approx 0.0610 \\ Sh_4(v) &\approx 0.1827 & Sh_5(v) &\approx 0.3789 \end{aligned}$$

The order of the variables, based on the relative importance metrics stayed the same. However direct comparison of the values cannot be done, because of the different distances. The two models and the results, show that the correlations between the predicted variables are misleading, when the predictors are correlated. The multicollinearity among the predictors effects the Shapley values.

This example also shows that the games in $\mathcal{G}_{LAD}^N$ are not necessarily superadditive or subadditive, this is also a property of the games in $\mathcal{G}_{LR}^N$ and $\mathcal{G}_{AR}^N$. A serious drawback of the previously introduced method, is that achieving the model itself is computationally demanding (McKean and Sievers, 1987), and computing the Shapley values – the relative importance measures – of models with high number of predictors is additionally demanding (Pintér, 2007).

3.3 Binary regression games

In the previously introduced regression game models, the predicted random variable was always a continuous random variable. In the following we introduce a new type of regression game, which has a categorical dependent random variable. The introduced model has a binary dependent variable with two possible outcomes. These binary dependent variable regression models are used in many areas such as bio-statistics (Woosung and Amid, 2005), mechanical engineering (Yan and Lee, 2004), marketing (Elrod and Keane, 1995) or portfolio management (Westgaard and der Wijst, 2001).

3.3.1 The class of binary regression games

Consider the following linear regression model:

$$\eta^\circ = \beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i \quad (3.7)$$

Let η° be the so-called latent dependent random variable (Maddala, 1999), and $\xi_i, i = 1, \dots, n$ are the set of predictor variables. We suppose that the random variables are known, and we do not have to estimate them. We define error variable ε as follows:

$$\varepsilon \equiv \eta^\circ - \beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i \quad (3.8)$$

Definition 3.23. *We suppose that we cannot observe η° but, we are able to observe η which is a binary variable. We define η as:*

$$\eta = \begin{cases} 1, & \text{if } \eta^\circ > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.9)$$

We assume that the error variable has a symmetric distribution (Wooldridge, 2012). Based on Definition 3.9 we can rewrite the conditional probability for η , for the previously fixed set of predictors:

$$P(\eta = 1) = P(\eta^\circ > 0)$$

The model defined by Definition 3.7 and the error variable of the model defined by Definition 3.8 allows us to rewrite the expression:

$$P(\eta = 1) = P\left(\beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i + \varepsilon > 0\right)$$

From this we obtain:

$$P(\eta = 1) = P\left(\varepsilon > -\left(\beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i\right)\right)$$

From the cumulative distribution function, and the assumption of symmetric error variable, it can be derived that:

$$\begin{aligned} P(\eta = 1) &= 1 - F\left(-\left(\beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i\right)\right) \\ P(\eta = 1) &= F\left(\beta_0 + \sum_{i=1}^n \beta_i \cdot \xi_i\right) \end{aligned} \quad (3.10)$$

Definition 3.24. Let $N = \{\xi_1, \dots, \xi_n\}$ be the players set of the n predictors.

In the following we assume that η has a binomial distribution and N is fixed, which means that the full model includes n predictor variables, and the empty model includes no predictor variables. Let us consider the following optimization problem, that can be rewritten with identity (3.10) as:

$$\begin{aligned} \mathcal{L} &= \prod P(\eta = 1) \cdot \prod P(\eta = 0) \rightarrow \max \\ \text{s.t. } &\beta_i \in \mathbb{R}, \quad i \in S \end{aligned} \quad (3.11)$$

Definition 3.25. Let η be the predicted variable, and $\xi_1, \dots, \xi_n$ be the predictor variables. For any $S \in \mathcal{P}(N)$, let $v(S)$ be the optimal objective function of (3.11).

The payoffs of the coalitions are defined by the goodness of fit (GOF) in the model. The measure of GOF in this model is the likelihood of the binary model. With logarithmic transformation of the objective function we could obtain the widely used log-likelihood value of the model. Because the logarithmic function is a monotone increasing function, this transformation does not affect the optimum.

Corollary 3.26. *Function v in Definition 3.25 is a game.*

We take the predictor variables as the players, and the subsets of predictor variables in the model as the coalitions. The GOF values give the payoffs of the coalitions of N .

Definition 3.27. *We call the games defined by Definitions 3.24 and 3.25 binary regression games. The class of binary regression games is denoted by $\mathcal{G}_{BR}^N$.*

Previously we only assumed, that the error variable in the latent variable model, has symmetric distribution and, this gives us the freedom to define different models. We will introduce two further binary models, that are commonly used in the econometric literature.

Definition 3.28. *If the error variable ε has a logistic distribution we call the regression models described by Equations (3.7),(3.8) and (3.9) logit regression models, the games defined by Definitions 3.24 and 3.25 logit regression games.*

Definition 3.29. *If the error variable ε has a normal distribution we call the regression models described by Equations (3.7),(3.8) and (3.9) probit regression models, the games defined by Definitions 3.24 and 3.25 probit regression games.*

Because of the plausible specification of the error variable, the probit model is used more frequently in econometrics, than the logit binary model. However, the logit model is also used widely in the literature, especially in financial econometrics e.g. mortgage securities (Kovács et al., 2011), and marketing research (Lipovetsky and Conklin, 2004).

Proposition 3.30. *The class of $\mathcal{G}_{BR}^N$ is a subset of the class of monotone games.*

An additional predictor variable in the model never decreases the likelihood of the model (Wooldridge, 2012), it means that the payoffs never decrease when a new player joins to the coalitions. This implies that the games in $\mathcal{G}_{BR}^N$ are monotone. Each type of binary regression models is defined with a particular error variable, this implies that straightforward comparison of the variables, from different models cannot be done. However, the comparison of relative importance orders of the variables can be relevant among the differently specified models.

3.3.2 Goodness of fit

Defining measure of GOF for binary regression models is complicated, and not straightforward. First without a binary model, that uses predictors at least 50% of the cases can be predicted correctly. This also means that metrics based on the number of cases predicted correctly are misleading and useless. Another problem with this method, is that the ratio

of correctly predicted cases always include an arbitrarily chosen cut value. The use of the objective function value is also problematic because the maximum of the function is zero. The logarithmic transformation of the likelihood values are not positive, so this rules out the immediate use of log-likelihood values. We introduce the pseudo- R^2 approach of McFadden (1974), which is quite universal for binary regression models.

Definition 3.31. Let $\mathcal{L}_0$ denote the log-likelihood of the empty model, which contains no predictors, we know the players set is empty.

Definition 3.32. Let $\mathcal{L}_1$ denote the log-likelihood of the actual model with a fixed set of $\{\xi_1, \dots, \xi_n\}$ predictors.

The measure of GOF is achieved by the McFadden pseudo- R^2 which can be computed as follows:

$$\text{McFadden pseudo-}R^2 = \frac{\mathcal{L}_0 - \mathcal{L}_1}{\mathcal{L}_0}$$

This measure of GOF has a value between 0 and 1. The perfect fit is achieved when the measure of GOF is 1, the 0 value shows us that the actual model's fit is not better than the empty model's fit.

Example 3.33. Consider the following binary regression model, where we observe the properties of some used cars (The database that is used can be found in Kane (2002)):

$$\eta^\circ = \beta_0 + \sum_{i=1}^5 \beta_i \cdot \xi_i$$

where η° is a latent variable, and we actually observe η which is binary variable. When the value of η is 1, it means that the car was involved in speeding in the last five years, otherwise was not. The predictor variables are the following:

ξ_1 :	Suggested original retail price in USA dollar	(Price)
ξ_2 :	Number of cylinders	(Cylinder)
ξ_3 :	Vehicle height in inch	(Height)
ξ_4 :	Vehicle weight in pound	(Weight)
ξ_5 :	Engine displacement in liter	(Displacement)

We define ε as the error variable like the following:

$$\varepsilon \equiv \eta - \beta_0 + \sum_{i=1}^5 \beta_i \cdot \xi_i$$

We assume, that the random variable ε has a logistic distribution. Let $N = \{\xi_1, \xi_2, \xi_3, \xi_4, \xi_5\}$ be the set of players, and let the McFadden pseudo- R^2 values of the models be the

payoff. The detailed table containing the payoffs of the coalitions is included in Section B.7 of the Appendix. Based on the payoffs, the Shapley values of the players can be computed:

$$\begin{aligned} Sh_1(v) &\approx 0.3382 & Sh_2(v) &\approx 0.0566 & Sh_3(v) &\approx 0.0481 \\ Sh_4(v) &\approx 0.0731 & Sh_5(v) &\approx 0.1189 \end{aligned}$$

The most important predictor is the cars suggested original retail price, it accounts for nearly 34% of GOF. The second in the relative importance order is the displacement of the engine. It should be noted, that the $Sh_5(v)$ is approximately 10 times greater than the $v(\{\xi_5\})$ payoff. This and the fact, that $Sh_1(v) < v(\{\xi_1\})$ shows us that the, predictor variables are strongly correlated, and collinearity is present in the logit regression model. The vehicle's weight accounts for nearly 7% percent of the model's distance from the empty model. The number of cylinders, and the vehicles height are the least important predictors in the logit model. Together these predictors account for near 10% percents of the models GOF. With a bar diagram we can show these relations:

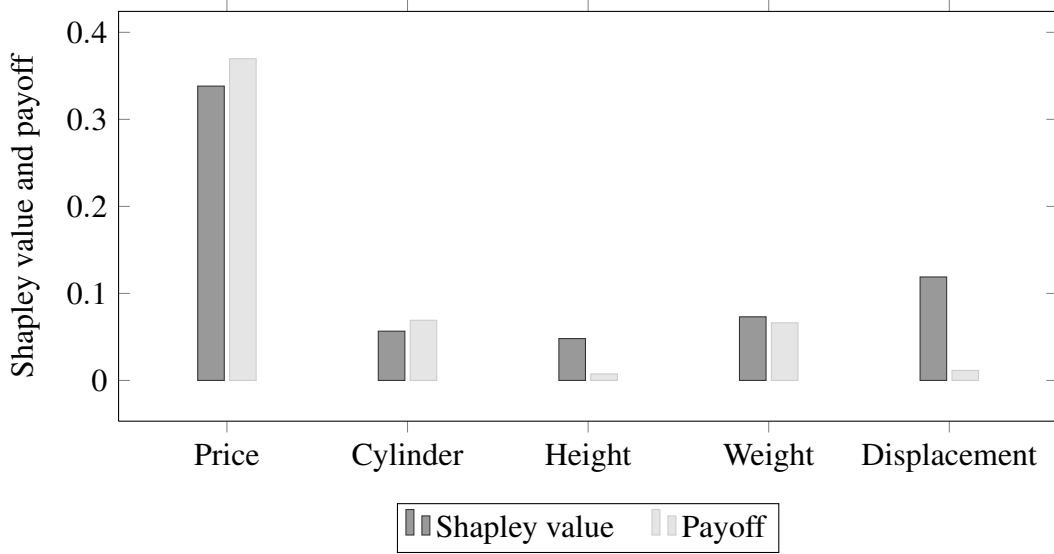


Figure 3.3: Shapley values and singleton coalition payoffs in Example 3.33

Claim 3.34. *A binary regression game might be an essential game.*

It can be shown, with Example 3.33, because the following relation stands between the payoffs of the coalitions and the payoff of the grand coalition:

$$\sum_{i=1}^5 v(\{\xi_i\}) < v(N)$$

This can be a property of games, that are in $\mathcal{G}_{LR}^N$ or in $\mathcal{G}_{AR}^N$. This also implies, that other solution concepts, that require the game to be essential, can be used in some cases

when the game is in $\mathcal{G}_{BR}^N$. Example 3.33 also shows that, games in $\mathcal{G}_{BR}^N$ are not necessarily superadditive or subadditive, this is also a property of the games in $\mathcal{G}_{LAD}^N$, $\mathcal{G}_{LR}^N$ and $\mathcal{G}_{AR}^N$.

Chapter 4

Summary

We have introduced the concept of transferable utility cooperative games, and we have highlighted some properties of these games with examples. We presented the Shapley value, a solution concept of cooperative game theory, and we showed some properties of the Shapley value. We introduced the class of linear regression games $\mathcal{G}_{LR}^N$, and we have applied the Shapley value as a relative importance measure. We made an argument based on the axiomatization that the Shapley value can be applied for relative importance ordering of predictor variables in regression models.

We broadened the concept of regression games, and resolved the restrictions of the original regression game approach, that was restricted to linear regression. We introduced three new classes of regression games: autoregressive games $\mathcal{G}_{AR}^N$, least absolute deviation games $\mathcal{G}_{LAD}^N$ and binary regression games $\mathcal{G}_{BR}^N$. Each of these classes of games is based on different optimization problems, and we defined the payoffs with these optimization problems. The newly introduced games are transferable utility cooperative games that are monotone, and might be subadditive or superadditive.

We have applied the Shapley value as a solution on these games. With the Shapley value we have shown, that the goodness of fit in the three models can be decomposed into relative importance shares. We have also showed with examples that the achieved relative importance measures can be used for relative importance ordering of the predictor variables. We also showed that Shapley value applied as a relative importance measure holds certain beneficial properties, that are useful in relative importance analysis and required by many authors according to Grömping (2007).

Our contributions to the application of regression games have shown, that the general model of regression games can be modified. The paper introduced regression games with known random variables. However in applied econometrics the random variables are estimated. It follows that the goodness of fit decompositions, and relative importance measures are also estimated. It is clear that an extension of the regression games to esti-

mated random variables can be done. The use of estimated variables would allow us the application of statistical hypothesis testing. The examples in Subsection 3.1.2 showed that the Shapley value can be a useful model selection tool. On each class of regression games introduced in the paper the Shapley value based model selection test can be applied.

Newly introduced regression games are only restricted to least absolute deviation games, binary regression games and autoregressive games. However, it is clear that each regression optimization problem corresponds to a class of regression games. If the game defined by the optimization problem is a transferable utility cooperative game, the Shapley value can be used for goodness of fit decomposition, and also for relative importance measuring. The introduction of non-linear regression functions e.g. exponential, logarithmic or trigonometric can be done too because basically the same problem has to be solved (non-linear least squares).

Based on the introduction of logit and probit regression games, another branch – likelihood based regression game models can be introduced like multinomial logit, ordered probit, autoregressive-moving average or Poisson-regression games. In many regression models the predictor variables and the parameters are in groups e.g. trigonometric time trend models or models with multiple slope. Cooperative game theory has solution concepts that can deal with such groups. This broadening of the regression games concept can be interesting also.

Bibliography

- Peter Bloomfield and William Steiger. Least Absolute Deviations Curve-Fitting. *Journal on Scientific Computing*, 1:290–301, 1980.
- Kani Chen, Zhiliang Ying, Hong Zhang, and Lincheng Zhao. Analysis of Least Absolute Deviation. *Biometrika*, 95:107–122, 2008.
- David Dickey and Wayne Fuller. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366):427–431, 1979.
- Terry Elrod and Michael Keane. A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data. *Journal of Marketing Research*, 32:1–16, 1995.
- Michael Feldman. The Proportional Value of a Cooperative Game. In *Manuscript for a contributed paper at the Econometric Society World Congress*, 2000.
- Ferenc Forgó, Miklós Pintér, András Simonovits, and Tamás Solymosi. Kooperatív játékelmélet, 2006. URL http://www.bke.hu/opkut/letoltheto_anyagok.html.
- Ulriche Grömping. Relative Importance for Linear Regression in R: The Package relaimpo., 2006.
- Ulriche Grömping. Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. *The American Statistician*, 61:139–146, 2007.
- Ottó Hajdu. Statisztikai adatok ökonometriai elemzése, Egyetemi jegyzet 1. rész, Keresztmetszeti modellek elmélete és alkalmazása, 2013a.
- Ottó Hajdu. Statisztikai adatok ökonometriai elemzése, Egyetemi jegyzet 2. rész, Idősori modellek elmélete és alkalmazása, 2013b.
- Edward James Hannan and Barry Quinn. The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society Series*, 41:190–195, 1979.
- Frank Huettner and Marco Sunder. Axiomatic Argument for Decomposing the Goodness of Fit According to Shapley and Owen values. *Electronic Journal of Statistics*, 6:1239–1250, 2012.
- John Kane. Econometrics: An Applied Approach - Cars Database, 2002. URL <http://www.oswego.edu/~kane/econometrics/cars.htm>.

- Erzsébet Kovács, Borbála Szüle, Vilmos Fliszár, and Péter Vékás. *Pénzügyi adatok statisztikai elemzése: Egyetemi tankönyv*. Tanszék Kft., 2011.
- Eugene Krause. *Taxicab Geometry*. Courier Dover, 1986.
- William Kruskal. Relative Importance by Averaging Over Orderings. *The American Statistician*, 41:6–10, 1987.
- Tarald Kvalseth. Cautionary Note about the Coefficient of Determination. *American Statistician*, 39:279–285, 1985.
- Stan Lipovetsky and Michael Conklin. Analysis of Regression in Game Theory Approach. *Applied Stochastic Models in Business and Industry*, 17:319–330, 2001.
- Stan Lipovetsky and Michael Conklin. Decision Making by Variable Contribution in Discriminant, Logit, and Regression Analyses. *International Journal of Information Technology and Decision Making*, 3:265–279, 2004.
- Stan Lipovetsky and Michael Conklin. Incremental Net Effects in Multiple Regression. *International Journal of Mathematical Education in Science and Technology*, 36(4):361–373, 2005.
- Kameswari (G.S.) Maddala. *Introduction to Econometrics*. Wiley and sons, 1999.
- Daniel McFadden. *Frontiers in Econometrics: Conditional Logit Analysis of Qualitative Choice Behavior*. Academic Press, 1974.
- Joseph McKean and Gerald Sievers. Coefficients of Determination for Least Absolute Deviation Analysis. *Statistics and Probability Letters*, 5:49–54, 1987.
- Anna Mikusheva. Time Series Analysis - Stationarity, Lag Operator, ARMA, and Covariance Structure. MIT Open Course Ware, 2007. URL <http://ocw.mit.edu>.
- Miklós Pintér. A regressziós játékok alkalmazása modellszelekcióra. 2006.
- Miklós Pintér. Regressziós játékok. *Sigma*, 38(4):131–148, 2007.
- Miklós Pintér. Regression Games. *Annals of Operations Research*, 186(1):263–274, 2011.
- Lloyd Shapley. A Value for n-person Games. *Contributions to the Theory of Games Volume II*, 28:307–317, 1953.
- John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- Sjur Westgaard and Nico Van der Wijst. Default Probabilities in a Corporate Bank Portfolio: A Logistic Model Approach. *European Journal of Operational Research*, 135:338–349, 2001.

Jeffrey Wooldridge. *Introductory Econometrics: A Modern Approach*. Cengage Learning, 2012.

Sohn Woosung and Ismail Amid. Regular Dental Visits and Dental Anxiety in an Adult Dentate Population. *Journal of the American Dental Association*, 136:58–67, 2005.

Jihong Yan and Jay Lee. Degradation Assessment and Fault Modes Classification Using Logistic Regression. *Journal of Manufacturing Science and Engineering*, 127:912–914, 2004.

Hobart Peyton Young. Monotonic Solutions of Cooperative Games. *International Journal of Game Theory*, 14:65–72, 1985.

Appendix A

Cooperative games

A.1 The *randlrg* MATLAB function

The function has only one independent variable - n , which is the number of predictor variables in the model, and the number of players in the TU game. We consider that the game is a linear regression game, denoted by $\mathcal{G}_{LR}^N$. The correlation matrix of the n predictors and the predicted variable will be an $(n+1) \times (n+1)$ symmetric positive definite matrix. The $\mathbf{A}$ matrix will be an $(n+1) \times (n+1)$ correlation generated randomly:

```
1 A = gallery('randcorr',n+1);
```

To obtain all of the goodness of fit values, we must recreate all of the correlation matrices resembling different linear regression models. Let $\mathbf{x}$ a binary row vector with n columns. The vector is a boolean and the non zero value means that a variable is in the certain regression model. For example the row vector $\mathbf{x} = [0\ 1\ 1]$ represents a set of predictor variables of $n = 3$, and in the model only the 2. and 3. variable is present. To calculate all of the goodness of fit values we must obtain all of these $\mathbf{x}$ boolean vectors, the matrix containing all of the vectors is denoted by $\mathbf{X}$.

```
1 x = [0; 1];
2 for i = 2:n,
3     s = size(x,1);
4     x = [[x zeros(s,1)]; [x ones(s,1)]];
5 end
```

From the matrix $\mathbf{X}$ we must create row vectors with length $n+1$ denoted by $\mathbf{k}$. We assume that $\mathbf{k}_{1,1} = 1$, and the remaining components are the respective rows of $\mathbf{X}$. In the previously mentioned example the vector $\mathbf{k}$ is the following $\mathbf{k} = [1\ 0\ 1\ 1]$. Based on the

different $\mathbf{k}$ vectors we can create logical matrices resembling the different models, based on the predictor variables. These logical matrices obtained like this: $\mathbf{B} = \mathbf{k}^\top \mathbf{k}$. In our example $\mathbf{B}$ is as follows:

$$\mathbf{B} = \mathbf{k}^\top \mathbf{k} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

With this $\mathbf{B}$ matrix of booleans we could easily choose the correlation values from the original correlation matrix $\mathbf{A}$. With reshaping the reduced correlation matrix we obtain the matrix $\mathbf{R}$. Lets imagine that in our example the $\mathbf{A}$ correlation matrix was the following:

$$\mathbf{A} = \begin{pmatrix} 1 & 0.7 & 0.4 & -0.5 \\ 0.7 & 1 & 0.3 & 0.8 \\ 0.4 & 0.3 & 1 & -0.2 \\ -0.5 & 0.8 & -0.2 & 1 \end{pmatrix}$$

Then the reduced $\mathbf{R}$ matrix based on $\mathbf{A}$ and $\mathbf{B}$ are as follows:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.4 & -0.5 \\ 0.4 & 1 & -0.2 \\ -0.5 & -0.2 & 1 \end{pmatrix}$$

A general form of $\mathbf{R}$ can be formalized like as it follows:

$$\mathbf{R} = \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,k} \\ c_{2,1} & c_{2,2} & \dots & c_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k,1} & c_{k,2} & \dots & c_{k,k} \end{pmatrix}$$

To calculate the multiple coefficient of determination we need the following column vectors from matrix $\mathbf{R}$:

$$\mathbf{c} = \begin{pmatrix} c_{2,1} \\ \vdots \\ c_{k,1} \end{pmatrix}$$

We also need the matrix denoted by $\mathbf{m}$, which can be obtained from $\mathbf{R}$:

$$\mathbf{m} = \begin{pmatrix} c_{2,2} & \dots & c_{2,k} \\ \vdots & \ddots & \vdots \\ c_{k,2} & \dots & c_{k,k} \end{pmatrix}$$

Based on all of these definitions, the multiple coefficient of determination (denoted by R^2) is calculated like as follows:

$$R^2 = \mathbf{c}^\top \mathbf{m}^{-1} \mathbf{c}$$

In our example with matrix **A** and **B**:

$$\mathbf{c} = \begin{pmatrix} 0.4 \\ -0.5 \end{pmatrix} \quad \mathbf{m} = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix}$$

Thus the R^2 value is the following:

$$R^2 = \mathbf{c}^\top \mathbf{m}^{-1} \mathbf{c} = 0.49$$

This is the MATLAB code sample that generates all of the goodness of fit values:

```
1 bound = 2^n;
2 for i = 1:bound
3   K = [1 x(i,:)];
4   B = K'*K;
5   C = logical(B);
6   Redvect = A(C);
7   b = length(Redvect)^0.5;
8   R = reshape(Redvect,b,b);
9   m = R(2:end,2:end);
10  c = R(2:end,1);
11  Rsq(i) = c'*inv(m)*c;
12 end
```

With the 2^n length $\mathbf{R}_{sq}$ vector, which includes all of the different goodness of fit values, we can generate a payoff vector v , which has a length $2^n - 1$.

```
1 v = Rsq(2:end);
```

Now we use the package MATTUG¹ based on the payoff vector v can calculate the Shapley values of every predictor:

```
1 Shapleyreg = ShapleyValue(v);
```

¹It is a free transferable utility game tool deigned for MATLAB and Octave

The complete MATLAB function:

```
1 function [Shapleyreg] = randlrg(n)
2 A = gallery('randcorr',n+1);
3 x = [0; 1];
4 for i = 2:n,
5 s = size(x,1);
6 x = [[x zeros(s,1)]; [x ones(s,1)]];
7 end
8 bound = 2^n;
9 for i = 1:bound
10 K = [1 x(i,:)];
11 B = K'*K;
12 C = logical(B);
13 Redvect = A(C);
14 b = length(Redvect)^0.5;
15 R = reshape(Redvect,b,b);
16 m = R(2:end,2:end);
17 c = R(2:end,1);
18 Rsq(i) = c'*inv(m)*c;
19 end
20 v = Rsq(2:end);
21 Shapleyreg = ShapleyValue(v);
22 end
```

A.2 The *reallrg* MATLAB function

The function only needs one variable to start. the **A** correlation matrix, which shows the different correlation between a set of predictors and a predicted variable. We assume that we will model a linear regression, and we can obtain the R^2 as a goodness of fit. The output is the Shapley value vector of the predictors.

```
1 function [Shapleyreg] = reallrg(A)
2 n = length(A)-1;
3 x = [0; 1];
4 for i = 2:n,
5 s = size(x,1);
6 x = [[x zeros(s,1)]; [x ones(s,1)]];
7 end
8 bound = 2^n;
9 for i = 1:bound
10 K = [1 x(i,:)];
11 B = K'*K;
12 C = logical(B);
13 Redvect = A(C);
14 b = length(Redvect)^0.5;
15 R = reshape(Redvect,b,b);
16 m = R(2:end,2:end);
17 c = R(2:end,1);
18 Rsq(i) = c'*inv(m)*c;
19 end
20 v = Rsq(2:end);
21 Shapleyreg = ShapleyValue(v);
22 end
```

Appendix B

Applications

B.1 Autoregressive simulation in Example 3.10

```
1 alpha=3; % Slope parameter
2 sigma=4; % Deviation of noise
3 mu=0; % Expected value of noise
4 phi1=0.5; % AR 1 parameter
5 phi4=0.4; %AR 4 parameter
6 Y(1:5)=30; % Expected value of the process
7 for i=6:1000000
8   Y(i)=alpha+phi1*Y(i-1)+phi4*Y(i-4)+normrnd(mu,sigma);
9 end
```

B.2 Autoregressive correlation matrices in Example 3.10

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \\ \eta_{-5} \end{matrix} \begin{pmatrix} 1 & 0.74 & 0.61 & 0.6 & 0.7 & 0.64 \\ 0.74 & 1 & 0.74 & 0.61 & 0.6 & 0.7 \\ 0.61 & 0.74 & 1 & 0.74 & 0.61 & 0.6 \\ 0.6 & 0.61 & 0.74 & 1 & 0.74 & 0.61 \\ 0.7 & 0.6 & 0.61 & 0.74 & 1 & 0.74 \\ 0.64 & 0.7 & 0.6 & 0.61 & 0.74 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-2} \\ \eta_{-4} \\ \eta_{-5} \end{matrix} \begin{pmatrix} 1 & 0.74 & 0.61 & 0.7 & 0.64 \\ 0.74 & 1 & 0.74 & 0.6 & 0.7 \\ 0.61 & 0.74 & 1 & 0.61 & 0.6 \\ 0.7 & 0.6 & 0.61 & 1 & 0.74 \\ 0.64 & 0.7 & 0.6 & 0.74 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-4} \\ \eta_{-5} \end{matrix} \begin{pmatrix} 1 & 0.74 & 0.7 & 0.64 \\ 0.74 & 1 & 0.6 & 0.7 \\ 0.7 & 0.6 & 1 & 0.74 \\ 0.64 & 0.7 & 0.74 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-4} \end{matrix} \begin{pmatrix} 1 & 0.74 & 0.7 \\ 0.74 & 1 & 0.6 \\ 0.7 & 0.6 & 1 \end{pmatrix}$$

B.3 Autoregressive simulation in Example 3.12

```

1 alpha=5; % Slope parameter
2 sigma=2; % Deviation of noise
3 mu=0; % Expected value of noise
4 phi2=0.5; % AR 2 parameter
5 phi3=0.05; %AR 3 parameter
6 phi4=0.4; %AR 4 parameter
7 Y(1:5)=100; % Expected value of the process
8 for i=6:1000000
9 Y(i)=alpha+phi2*Y(i-2)+phi3*Y(i-3)+phi4*Y(i-4)+normrnd(mu,sigma);
10 end

```

B.4 Autoregressive correlation matrices in Example 3.12

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \\ \eta_{-5} \\ \eta_{-6} \\ \eta_{-7} \end{matrix} \begin{pmatrix} 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 & 0.81 & 0.51 \\ 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 & 0.81 \\ 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 \\ 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 \\ 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 \\ 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 \\ 0.81 & 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 \\ 0.51 & 0.81 & 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \\ \eta_{-5} \\ \eta_{-6} \end{matrix} \begin{pmatrix} 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 & 0.81 \\ 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.51 \\ 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.87 \\ 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.52 \\ 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.88 \\ 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.51 \\ 0.81 & 0.51 & 0.87 & 0.52 & 0.88 & 0.51 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-1} \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \\ \eta_{-6} \end{matrix} \begin{pmatrix} 1 & 0.51 & 0.88 & 0.52 & 0.87 & 0.81 \\ 0.51 & 1 & 0.51 & 0.88 & 0.52 & 0.51 \\ 0.88 & 0.51 & 1 & 0.51 & 0.88 & 0.87 \\ 0.52 & 0.88 & 0.51 & 1 & 0.51 & 0.52 \\ 0.87 & 0.52 & 0.88 & 0.51 & 1 & 0.88 \\ 0.81 & 0.51 & 0.87 & 0.52 & 0.88 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \\ \eta_{-6} \end{matrix} \begin{pmatrix} 1 & 0.88 & 0.52 & 0.87 & 0.81 \\ 0.88 & 1 & 0.51 & 0.88 & 0.87 \\ 0.52 & 0.51 & 1 & 0.51 & 0.52 \\ 0.87 & 0.88 & 0.51 & 1 & 0.88 \\ 0.81 & 0.87 & 0.52 & 0.88 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{matrix} \eta \\ \eta_{-2} \\ \eta_{-3} \\ \eta_{-4} \end{matrix} \begin{pmatrix} 1 & 0.88 & 0.52 & 0.87 \\ 0.88 & 1 & 0.51 & 0.88 \\ 0.52 & 0.51 & 1 & 0.51 \\ 0.87 & 0.88 & 0.51 & 1 \end{pmatrix}$$

B.5 LAD game payoffs in Example 3.22

Coalition	City	Length	Height	Weight	Power	GOF
$\emptyset$	No	No	No	No	No	0
$\{\xi_1\}$	Yes	No	No	No	No	0.2514
$\{\xi_2\}$	No	Yes	No	No	No	0.0759
$\{\xi_3\}$	No	No	Yes	No	No	0.0167
$\{\xi_4\}$	No	No	No	Yes	No	0.2895
$\{\xi_5\}$	No	No	No	No	Yes	0.4730
$\{\xi_1, \xi_2\}$	Yes	Yes	No	No	No	0.2518
$\{\xi_1, \xi_3\}$	Yes	No	Yes	No	No	0.2784
$\{\xi_1, \xi_4\}$	Yes	No	No	Yes	No	0.2920
$\{\xi_1, \xi_5\}$	Yes	No	No	No	Yes	0.4744
$\{\xi_2, \xi_3\}$	No	Yes	Yes	No	No	0.0913
$\{\xi_2, \xi_4\}$	No	Yes	No	Yes	No	0.2940
$\{\xi_2, \xi_5\}$	No	Yes	No	No	Yes	0.4753
$\{\xi_3, \xi_4\}$	No	No	Yes	Yes	No	0.3915
$\{\xi_3, \xi_5\}$	No	No	Yes	No	Yes	0.4733
$\{\xi_4, \xi_5\}$	No	No	No	Yes	Yes	0.4894
$\{\xi_1, \xi_2, \xi_3\}$	Yes	Yes	Yes	No	No	0.2806
$\{\xi_1, \xi_2, \xi_4\}$	Yes	Yes	No	Yes	No	0.2967
$\{\xi_1, \xi_2, \xi_5\}$	Yes	Yes	No	No	Yes	0.4769
$\{\xi_1, \xi_3, \xi_4\}$	Yes	No	Yes	Yes	No	0.3924
$\{\xi_1, \xi_3, \xi_5\}$	Yes	No	Yes	No	Yes	0.4752
$\{\xi_1, \xi_4, \xi_5\}$	Yes	No	No	Yes	Yes	0.4893
$\{\xi_2, \xi_3, \xi_4\}$	No	Yes	Yes	Yes	No	0.5583
$\{\xi_2, \xi_3, \xi_5\}$	No	Yes	Yes	No	Yes	0.4754
$\{\xi_2, \xi_4, \xi_5\}$	No	Yes	No	Yes	Yes	0.4956
$\{\xi_3, \xi_4, \xi_5\}$	No	No	Yes	Yes	Yes	0.5051
$\{\xi_2, \xi_3, \xi_4, \xi_5\}$	No	Yes	Yes	Yes	Yes	0.5535
$\{\xi_1, \xi_3, \xi_4, \xi_5\}$	Yes	No	Yes	Yes	Yes	0.5084
$\{\xi_1, \xi_2, \xi_4, \xi_5\}$	Yes	Yes	No	Yes	Yes	0.5046
$\{\xi_1, \xi_2, \xi_3, \xi_5\}$	Yes	Yes	Yes	No	Yes	0.4780
$\{\xi_1, \xi_2, \xi_3, \xi_4\}$	Yes	Yes	Yes	Yes	No	0.4597
$\{\xi_1, \xi_2, \xi_3, \xi_4, \xi_5\}$	Yes	Yes	Yes	Yes	Yes	0.5653

Table B.1: LAD game payoffs of the coalitions in Example 3.22

B.6 Regression game payoffs in Example 3.22

Coalition	City	Length	Height	Weight	Power	GOF
$\emptyset$	No	No	No	No	No	0
$\{\xi_1\}$	Yes	No	No	No	No	0.3613
$\{\xi_2\}$	No	Yes	No	No	No	0.1091
$\{\xi_3\}$	No	No	Yes	No	No	0.0057
$\{\xi_4\}$	No	No	No	Yes	No	0.3635
$\{\xi_5\}$	No	No	No	No	Yes	0.7064
$\{\xi_1, \xi_2\}$	Yes	Yes	No	No	No	0.3640
$\{\xi_1, \xi_3\}$	Yes	No	Yes	No	No	0.4092
$\{\xi_1, \xi_4\}$	Yes	No	No	Yes	No	0.4022
$\{\xi_1, \xi_5\}$	Yes	No	No	No	Yes	0.4744
$\{\xi_2, \xi_3\}$	No	Yes	Yes	No	No	0.1091
$\{\xi_2, \xi_4\}$	No	Yes	No	Yes	No	0.7070
$\{\xi_2, \xi_5\}$	No	Yes	No	No	Yes	0.7094
$\{\xi_3, \xi_4\}$	No	No	Yes	Yes	No	0.5910
$\{\xi_3, \xi_5\}$	No	No	Yes	No	Yes	0.7065
$\{\xi_4, \xi_5\}$	No	No	No	Yes	Yes	0.7185
$\{\xi_1, \xi_2, \xi_3\}$	Yes	Yes	Yes	No	No	0.4118
$\{\xi_1, \xi_2, \xi_4\}$	Yes	Yes	No	Yes	No	0.7071
$\{\xi_1, \xi_2, \xi_5\}$	Yes	Yes	No	No	Yes	0.7110
$\{\xi_1, \xi_3, \xi_4\}$	Yes	No	Yes	Yes	No	0.5977
$\{\xi_1, \xi_3, \xi_5\}$	Yes	No	Yes	No	Yes	0.7071
$\{\xi_1, \xi_4, \xi_5\}$	Yes	No	No	Yes	Yes	0.7226
$\{\xi_2, \xi_3, \xi_4\}$	No	Yes	Yes	Yes	No	0.6734
$\{\xi_2, \xi_3, \xi_5\}$	No	Yes	Yes	No	Yes	0.7071
$\{\xi_2, \xi_4, \xi_5\}$	No	Yes	No	Yes	Yes	0.7347
$\{\xi_3, \xi_4, \xi_5\}$	No	No	Yes	Yes	Yes	0.7389
$\{\xi_2, \xi_3, \xi_4, \xi_5\}$	No	Yes	Yes	Yes	Yes	0.7840
$\{\xi_1, \xi_3, \xi_4, \xi_5\}$	Yes	No	Yes	Yes	Yes	0.7426
$\{\xi_1, \xi_2, \xi_4, \xi_5\}$	Yes	Yes	No	Yes	Yes	0.7412
$\{\xi_1, \xi_2, \xi_3, \xi_5\}$	Yes	Yes	Yes	No	Yes	0.7110
$\{\xi_1, \xi_2, \xi_3, \xi_4\}$	Yes	Yes	Yes	Yes	No	0.6741
$\{\xi_1, \xi_2, \xi_3, \xi_4, \xi_5\}$	Yes	Yes	Yes	Yes	Yes	0.7921

Table B.2: Regression game payoffs of the coalitions in Example 3.22

B.7 Logit regression game payoffs in Example 3.33

Coalition	Price	Cylinder	Height	Weight	Displacement	GOF
$\emptyset$	No	No	No	No	No	0
$\{\xi_1\}$	Yes	No	No	No	No	0.3697
$\{\xi_2\}$	No	Yes	No	No	No	0.0691
$\{\xi_3\}$	No	No	Yes	No	No	0.0075
$\{\xi_4\}$	No	No	No	Yes	No	0.0662
$\{\xi_5\}$	No	No	No	No	Yes	0.0115
$\{\xi_1, \xi_2\}$	Yes	Yes	No	No	No	0.4395
$\{\xi_1, \xi_3\}$	Yes	No	Yes	No	No	0.3967
$\{\xi_1, \xi_4\}$	Yes	No	No	Yes	No	0.3942
$\{\xi_1, \xi_5\}$	Yes	No	No	No	Yes	0.5692
$\{\xi_2, \xi_3\}$	No	Yes	Yes	No	No	0.0903
$\{\xi_2, \xi_4\}$	No	Yes	No	Yes	No	0.0803
$\{\xi_2, \xi_5\}$	No	Yes	No	No	Yes	0.1698
$\{\xi_3, \xi_4\}$	No	No	Yes	Yes	No	0.2378
$\{\xi_3, \xi_5\}$	No	No	Yes	No	Yes	0.0255
$\{\xi_4, \xi_5\}$	No	No	No	Yes	Yes	0.0802
$\{\xi_1, \xi_2, \xi_3\}$	Yes	Yes	Yes	No	No	0.4479
$\{\xi_1, \xi_2, \xi_4\}$	Yes	Yes	No	Yes	No	0.4403
$\{\xi_1, \xi_2, \xi_5\}$	Yes	Yes	No	No	Yes	0.5991
$\{\xi_1, \xi_3, \xi_4\}$	Yes	No	Yes	Yes	No	0.3974
$\{\xi_1, \xi_3, \xi_5\}$	Yes	No	Yes	No	Yes	0.5694
$\{\xi_1, \xi_4, \xi_5\}$	Yes	No	No	Yes	Yes	0.5849
$\{\xi_2, \xi_3, \xi_4\}$	No	Yes	Yes	Yes	No	0.2684
$\{\xi_2, \xi_3, \xi_5\}$	No	Yes	Yes	No	Yes	0.1810
$\{\xi_2, \xi_4, \xi_5\}$	No	Yes	No	Yes	Yes	0.2168
$\{\xi_3, \xi_4, \xi_5\}$	No	No	Yes	Yes	Yes	0.3868
$\{\xi_2, \xi_3, \xi_4, \xi_5\}$	No	Yes	Yes	Yes	Yes	0.4266
$\{\xi_1, \xi_3, \xi_4, \xi_5\}$	Yes	No	Yes	Yes	Yes	0.6170
$\{\xi_1, \xi_2, \xi_4, \xi_5\}$	Yes	Yes	No	Yes	Yes	0.6121
$\{\xi_1, \xi_2, \xi_3, \xi_5\}$	Yes	Yes	Yes	No	Yes	0.5994
$\{\xi_1, \xi_2, \xi_3, \xi_4\}$	Yes	Yes	Yes	Yes	No	0.4602
$\{\xi_1, \xi_2, \xi_3, \xi_4, \xi_5\}$	Yes	Yes	Yes	Yes	Yes	0.6349

Table B.3: Logit regression game payoffs of the coalitions in Example 3.33