APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind., 2001; **17**:319–330 (DOI: 10.1002/asmb.446)

Analysis of regression in game theory approach

Stan Lipovetsky*,† and Michael Conklin

Custom Research Inc., 8401 Golden Valley Road, Minneapolis, MN 55427, U.S.A.

SUMMARY

Working with multiple regression analysis a researcher usually wants to know a comparative importance of predictors in the model. However, the analysis can be made difficult because of multicollinearity among regressors, which produces biased coefficients and negative inputs to multiple determination from presumably useful regressors. To solve this problem we apply a tool from the co-operative games theory, the Shapley Value imputation. We demonstrate the theoretical and practical advantages of the Shapley Value and show that it provides consistent results in the presence of multicollinearity. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: co-operative games; Shapley Value; multicollinearity; regressors net effects

1. INTRODUCTION

In this work we consider a common problem often faced by researchers and practitioners applying regression modelling—estimation of relative importance, or shares of influence that independent variables contribute to the model. Widely used statistical measures of predictors' importance include coefficients of regression, their t-statistics, partial F-statistics, p-values, inputs to multiple determination R^2 from net effects of each variable. For real data the variables are correlated, often highly correlated, and could be stochastically dependent in subsets of several variables, or multicollinear. Multicollinearity is not a problem for prediction by regression, but it has several detrimental effects in the analysis of the regressors' influence on the criterion variable. Such effects are as follows: parameter estimates can fluctuate wildly with a negligible change in the sample, they can have signs opposite to signs of easily understood pair-wise correlations, and theoretically important variables can have insignificant coefficients. Multicollinearity causes a reduction in statistical power, or the ability of statistical tests to detect true differences in the population. This leads to wider confidence intervals around the coefficients implying that they could be incorrectly identified as being insignificant, and the ability to determine the difference between parameters is degraded [1].

* Correspondence to: Stan Lipovetsky, Custom Research Inc., 8401 Golden Valley Road, Minneapolis, MN 55427, U.S.A.
† E-mail: lipovetsky@customresearch.com

Published online 1 October 2001
Received 17 July 2000
Revised 26 April 2001

Many different techniques are used in applied regression analysis to evaluate the relative importance of the predictors. One particularly useful technique is a decomposition of the coefficient of multiple determination into direct, indirect and net effects associated with each variable [2]. The net effect of a predictor is a combination of the direct (as measured by its regression coefficient squared) and the indirect effects (measured by the combination of its correlations with other variables). The net effects have the nice property of summing to the total coefficient of multiple determination R^2 of the model. They explicitly take into account the correlations that predictor variables have with each other. However, the net effect values themselves are influenced by the collinear redundancy in the data so that the estimated net effects can be negative, that is difficult to interpret. On the other hand, even in presence of multicollinearity, it is often desirable to keep available variables in the model and to estimate comparative importance of their relation to the dependent variable. It makes sense because all variables do not represent each other exactly, rather each of them plays its own specific role in fitting and understanding behaviour of the dependent variable.

Another important issue is that in most real-world situations, researchers and practitioners have neither a comprehensive theory, nor a complete control over all variables that could describe numerous specific features of a complex object or process (see, for example, References [3,4]). When a model does not incorporate some variables that correlate with included variables, regression estimates are biased and inconsistent (see References [5, pp. 334–350; 6, pp. 24–25]). However, in the absence of knowing what variables are necessary for the equation, we can consider different models, even limiting consideration to linear regressions. Also some kind of averaging of the estimated characteristics over all the possible models can be applied.

Various techniques were elaborated for choosing and averaging among the regression models to find the best subset [7–10]. However, given the power reduction caused by multicollinearity it is difficult to be sure that the best models found by such search are really superior to the many other models that are rejected. Many techniques have also been proposed for combining alternative models. Bagging, or bootstrap aggregation [11], boosting [12] and bundling [13] are just some of the techniques that have been shown to improve predictions over choosing a single model. The disadvantage in these techniques is that inference about specific variables is difficult.

In bootstrapping experiments we can demonstrate that a model with typical levels of multicollinearity could consist of spurious coefficients corresponding to a random distribution around zero. In the presence of multicollinearity, the standard error of every coefficient can be several times more and at least not less than the coefficient itself. This means that an actually arbitrary decision could be made based on the analysis of the regression coefficients or their shares of influence.

Thus, we need a decision tool that can produce clear results for estimation of regressors even if they are collinear, and when there actually are many possible models by various subsets of the predictors. The appropriate tool we find in the co-operative game theory. We can think of the particular model as a way of building coalitions among players (predictor variables) to maximize the total value (quality of fitting). In the field of co-operative games a useful analysis and decision tool is the Shapley Value imputation [14–17]. Results of the regression calculations can serve as the initial data to finding Shapley Value that in its turn reshapes the regression net effects and coefficients. We proved this technique to be useful during several years of solving various complicated problems in the marketing research field [18–23]. Of course, some other game-theoretical techniques can be used in statistical decisions under consideration, but we prefer the Shapley Value imputation because it is not an heuristic procedure, it was derived as an axiomatic approach, and it produces a unique solution satisfying general requirements of Nash equilibrium [14,15].

This paper is organized as following. In Section 2 we describe how to estimate the predictors importance, and in Section 3 we consider Shapley Value analysis. Its application to net effects and a possibility to adjust regression coefficients is considered in Section 4. Numerical example of Shapley Value in regression analysis with bootstrapping estimation is given in Section 5. In Section 6 we summarize results of Shapley Value application to analysing regression models.

2. PREDICTORS' CONTRIBUTION IN REGRESSION

Let us consider briefly some properties of linear regression that will be used in further analysis. A multiple regression model is

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_n x_{ni} + e_i \tag{1}$$

where y and all x's are standardized variables (centred and normalized by standard deviations), e denotes normal random error and i is a number of observation ($i = 1, 2, \ldots, N$). Coefficients of the regression can be found in the least-squares (LS) approach that corresponds to minimization of the objective

$$S^2 = e'e = (y - Xb)'(y - Xb) = 1 - 2b'r + b'Cb \tag{2}$$

where y and e are the Nth-order vector-columns of the dependent variable and error term, respectively, X is the matrix of $N \times n$ order of standardized independent variables, b is the nth-order vector-column of LS estimates for beta-coefficients of the regression (1), $C = X'X$ is the nth-order matrix of sample correlations r_{jk} among all pairs of variables x_j and x_k, vector-column $r = X'y$ of size n defines sample correlations r_j between each x_j and the dependent variable y and prime denotes transposition. Minimizing (2) yields the normal system of equations

$$Cb = r \tag{3}$$

and solution of system (3) is

$$b = C^{-1} r \tag{4}$$

where C^{-1} is the inverted matrix of correlations.

Minimizing deviations (2) is equivalent to maximizing of the regression quality estimated by the coefficient of multiple determination R^2 (this problem in the context of canonical analysis is discussed in References [24–26]), that could be presented as

$$R^2 = 1 - S^2 = 2b'r - b'Cb = \sum_j b_j\,(2r - Cb)_j \tag{5}$$

and reduced to the same solution (3) and (4). Substituting (3) into (5), we find the maximum of the multiple determination as the scalar product of vectors b and r:

$$R^2 = b'r = \sum_j b_j r_j \tag{6}$$

Items of the total R^2 (6) define the so-called net effects (NEF) of each jth regressor

$$\mathrm{NEF}_j = b_j r_j \tag{7}$$

Multicollinearity can change a sign of b_j in the multiple regression to opposite in comparison with the pairwise regression y by x_j (pair regression coefficient coincides with the correlation r_j), then jth net effect becomes negative.

The values of multiple determination (6) and net effects (7) are widely used in practice of regression modelling, although there are theoretically more reasonable indices for evaluation of predictor importance. For example, the share of x_j could be defined by the square of the partial correlation between y and x_j with fixed other x's (see References [27,28]):

$$R_{yj}^2 = (R^2 - R_{-j}^2)/(1 - R_{-j}^2) \tag{8}$$

where R^2 denotes multiple determination in model (1) with all n predictors including x_j, and R_{-j}^2 denotes multiple determination in the model with $n-1$ predictors without x_j.

Another measure of relative importance of predictors is considered in References [29, 30]. It consists in evaluation of usefulness of each regressor via the increment of multiple determination R^2 of the model with this particular x_j in the set of regressors in comparison with the model without x_j—i.e. just the numerator in (8):

$$U_j = R^2 - R_{-j}^2 \tag{9}$$

Coefficient (8) corresponds to a relative value of measure (9), and it reduces to (9) when $R_{-j}^2 \ll 1$. The characteristic of usefulness (9) can be represented as the scalar product

$$U_j = (b - b^*)'r \tag{10}$$

where b and r are the same vectors as in (6), and b^* is a vector of the nth order with zero in the position of the variable x_j and all the other elements equal to the coefficients in the regression model without x_j. It is possible to represent (10) as follows [30]:

$$U_j = b_j^2\ (1 - R_{j,\mathrm{others}}^2) \tag{11}$$

where b_j is the jth coefficient of regression (1) by all n variables, and $R_{j,\mathrm{others}}^2$ equals to the multiple determination in the regression of x_j by all the others $n-1$ predictors. Recognizing the second term in (11) as the jth variance inflation factor [31,32] we can rewrite (11) in a simple form

$$U_j = b_j^2 / C_{jj}^{-1} \tag{12}$$

where in denominator we have the jth diagonal element of the inverted correlation matrix C^{-1} (4) of all the x's.

In the presence of highly correlated variables among the x's, the matrix C (3) degenerates to the ill-conditioned matrix, and solution (4) produces poor results both for coefficients of regression and for net effects. In the limit of very high multicollinearity among the x's solution (4) simply does not exist, and the techniques of the ridge regression analysis can be recommended (for example, References [9,10,32]). Let us consider now new possibilities that the Shapley Value analysis suggests for estimation of regressors' shares in their mutual influence on the dependent variable, and for evaluation of the coefficients in the regression model.

3. SHAPLEY VALUE IMPUTATION

The Shapley Value, hereafter referred to as SV, was developed to evaluate an ordering of the worth of players in a multiplayer co-operative game. The key to understanding its utility is that it represents the worth of each player over all possible combinations of players. Extending this to the problem of comparative usefulness of regressors, the SV assigns a value for each predictor calculated over all possible combinations of predictors in regressions.

The SV approach to the problem provides a solution that is closer to the actual modeling for any complex process or object, because it compares and averages over all possible subsets of predictors in the model. This is an advantage of the SV solution because by comparing across all possible models it includes the possibility of competitive influence of any subsets of predictors in the analysis.

The Shapley Value is defined as each jth participant's input to a coalition

$$S_j = \sum_{\text{all } M} \gamma_n(M)\, [\upsilon\,(M \cup \{j\}) - \upsilon\,(M)] \tag{13}$$

with weights of proportions to enter into a coalition M defined as

$$\gamma_n(M) = m!(n-m-1)\,!/n! \tag{14}$$

In (13) and (14), n is the total number of all the participants, m is the number of participants in the Mth coalition, and υ ( ) is the characteristic function used for estimation of utility for each coalition. By $M \cup \{j\}$ a set of participants which includes the jth participant is denoted, when M means a coalition without the jth participant. In our case, the participants of the coalition game are predictors incorporated into the regression model.

Regression output supplies us with R^2 values, or per cent of explained variability reached for each set of variables in regression modelling. For ease of exposition, let us use notations A, B and C, etc. for variables x_1, x_2 and x_3, etc., respectively. Then R^2_{ABC}, for example, defines the multiple determination in model (1) with the regressors A, B and C (or, the same, x_1, x_2 and x_3).

We define characteristic function υ (13) via these R^2 values estimated by the results of regression modelling. Let us construct, for the example of $n = 5$, the characteristic function for the variable A (where there are other variables B, C, D and E):

$$\upsilon\,(0) = 0, \quad \upsilon\,(A) = R^2_A, \quad \upsilon\,(AB) = R^2_{AB}, \ldots, \upsilon\,(ABCDE) = R^2_{ABCDE}, \tag{15}$$

where all the right-hand-side values are estimations of the multiple determination coefficients in different regressions containing the regressor A. Substituting characteristic function (15) into SV expression (13), we can see that each item in brackets (13) actually coincides with the usefulness defined in (9). That means that SV for a predictor A is a measure of its usefulness averaged by all the models that contain this regressor A.

Weights of imputation (14) for $n = 5$ are

$$\gamma(0) = \gamma(4) = 0.20, \quad \gamma(1) = \gamma(3) = 0.05, \quad \gamma(2) = 0.033 \tag{16}$$

Then the SV (13) for the variable A can be written explicitly as

$$\begin{aligned} S_A = 0.2(U_A) &+ 0.05(U_{AB} + U_{AC} + U_{AD} + U_{AE}) \\ &+ 0.033(U_{ABC} + U_{ABD} + U_{ABE} + U_{ACD} + U_{ACE} + U_{ADE}) \\ &+ 0.05(U_{ABCD} + U_{ABCE} + U_{ACDE} + U_{ABDE}) + 0.2(U_{ABCDE}) \end{aligned} \tag{17}$$

where the values of usefulness (9) for our sets of regressors are as follows:

$$\begin{aligned} U_A = R_A^2, \quad U_{AB} = R_{AB}^2 - R_B^2, \ldots, U_{ABC} = R_{ABC}^2 - R_{BC}^2, \ldots \\ U_{ABCD} = R_{ABCD}^2 - R_{BCD}^2, \ldots, U_{ABCDE} = R_{ABCDE}^2 - R_{BCDE}^2 \end{aligned} \tag{18}$$

The items in sum (17) correspond to usefulness' margins from the variable A to all the coalitions, and the Shapley Value imputation S_A corresponds to the mean margin of the variable A, estimated by averaging over its possible participation in all coalitions. Similar formulas are used for each of the other variables B, C, D and E, and their Shapley Values (13) define margins from each of these regressors. The total of the margins from all the variables equals the maximum value of R^2 in the model with all the regressors together, that is (due to (15))

$$\sum_k^n S_k = \upsilon\,(\text{all}) = R_{ABCDE}^2 \tag{19}$$

Thus, the SV are shares of the total R^2 and they define importance of each regressor in the model.

Regrouping the items in (17) with help of (18), we represent the Shapley Value imputation formula in the following form:

$$\begin{aligned} S_A = (R_A^2 - \bar{R}_1^2)/(n-1) + (\bar{R}_{A*}^2 - \bar{R}_2^2)/(n-2) + (\bar{R}_{A**}^2 - \bar{R}_3^2)/(n-3) \\ + \cdots + (\bar{R}_{A**}^2 - \bar{R}_{n-1}^2)/(n-(n-1)) + R_{AB\ldots N}^2/n \end{aligned} \tag{20}$$

In the first item of sum (20) we see a difference of R_A^2 for the model with one regressor A and mean value of $\bar{R}_1^2$ (marked by bar over R^2) for all the models with just one regressor (marked by sub-index 1). In the second item of sum (20) we see is a difference between mean $\bar{R}_{A*}^2$ for all the models with two regressors one of which is A (marked by sub-index $A*$ with asterisk denoting any other variable x), and mean $\bar{R}_2^2$ for all the models with any two regressors (marked by sub-index 2), etc. The last item represents a share that the regressor A has in the total R^2 of model (1) with all the x's together.

4. SV NET EFFECTS AND COEFFICIENTS OF ADJUSTED REGRESSION

Suppose, we found SV (20) for each regressor, with their total equals multiple determination (19) for model (1) with all the variables. These Shapley Values (19) are nothing else but estimations of the net effects (7) obtained via averaging by all possible models in the co-operative game approach. Returning from lettered indices to the index j, let us denote by SV_j the net effects estimated by Shapley Value imputation for each regressor. Then in place of (6) and (7) for regular net effects for multiple regression (1) we can write decomposition of the multiple determination by the net effects estimated as Shapley Values

$$R^2 = \sum_j SV_j \tag{21}$$

Each item in (21) is a very stable estimate of net effect because Shapley Value is an average across all possible linear models with different subsets of the regressors. Thus, it is not so volatile as regular net effect and is not prone to multicollinearity distortion. In comparison to net effects (7), SV net effects (21) are always positive, so they are interpretable and suggest an easy way of graphical (pie-charts) presentation of regressors' shares in their contribution to explanation of the behaviour of dependent variable

$$\sum_j (\mathrm{SV}_j/R^2) = 1 \tag{22}$$

Let us consider a simple possibility of estimating which of the shares (22) are statistically significant, or which of the regressors are important in their contribution to R^2. Suppose we take level of significance α (for example, 5 per cent) for checking the difference of the coefficient R^2 (21) from zero. In the assumption of independence of the items in multiple comparison and for equal confidential probability $1-\gamma$ for each of them, we can obtain the joint probability $1-\alpha$ as the product of the probabilities of all the items $(1-\gamma)^n$. Then the so-called Bonferroni confidential interval [33] for each item among n of them can be defined as

$$\gamma = 1-(1-\alpha)^{1/n} \tag{23}$$

that for $n > 1$ is always less than α level. Standard deviation σ_R for the coefficient of multiple regression R (square root of multiple determination R^2) equals

$$\sigma_R = \sqrt{(1-R^2)/(N-n-1)} \tag{24a}$$

where $N - n - 1$ is the number of degrees of freedom. Let us denote sample t-value for R as $t_R = R/\sigma_R$. On the level of significance γ (23) the interval estimates for R are

$$R_{\pm} = R \pm t_{\gamma/2}\sigma_R \tag{24b}$$

with two-tailed t-statistics $t_{\gamma/2}$. Then the relative squared deviation is

$$\delta^2 = ((R_{\pm} - R)/R)^2 = t^2_{\gamma/2}\sigma^2_R/R^2 = t^2_{\gamma/2}/t^2_R \tag{25}$$

that equals the ratio of critical and empirical squared t-values. The shares of net effects (22) higher than the threshold (25), $\mathrm{SV}_j/R^2 > \delta^2$, can be considered as important (significantly different from zero), and those $\mathrm{SV}_j/R^2 < \delta^2$ correspond to the variables that can be neglected.

Let us consider adjusting regression coefficients by SV net effects. Using calculated SV_j, we can rewrite relation (7) for net effect as $\mathrm{SV}_j = a_j r_j$ where by a_j we denote unknown parameters of an adjusted regression with such a property: product of each coefficient a_j with correlations r_j yields Shapley Value net effect. Then solving these simple equations we have:

$$a_j = \mathrm{SV}_j/r_j. \tag{26}$$

Coefficients (26) are obtained via definition (7) for net effects. However, this definition (7) corresponds to items in expression (6) for multiple determination obtained as a result of substitution of the least-squares normal system of Equations (3) into the general objective for R^2 (5). To re-estimate coefficients of the regression with the obtained Shapley Values, we suggest to use not a simple relation $\mathrm{SV}_j = a_j r_j$ but a more complicated expression for net effect defined as the items in sum (5) with the coefficients a_j. So we can write equations for finding coefficients of regression adjusted by Shapley Values as follows:

$$a_j\,(2r - Ca)_j = \mathrm{SV}_j \tag{27}$$

For already found Shapley Values, relations (27) present a system of n quadratic equations. This system can be solved for a_j by a non-linear minimizing of the objective

$$F = \sum_j\,[\mathrm{SV}_j - a_j(2r - Ca)_j]^2 = \sum_j\,(\mathrm{SV}_j - 2a_j r_j + a_j \sum_k r_{jk}a_k)^2 \tag{28}$$

Coefficients (26) can be used as the initial approximation $a_j^{(0)}$ in the minimizing procedure (28). Parameters a_j obtained in (28) are coefficients of the adjusted regression estimated via Shapley Values. These coefficients are not prone to distortion from multicollinearity, and have interpretable signs and values.

Let us consider a convenient characteristic of the difference between two solutions for regression coefficients. It can be constructed as a ratio of the residual sum of squares for a non-least-squares regression to the residual sum of squares of the least-squares regression (see References [31,34]). Denote residual sum of squares in model (1) as $S^2(a)$ for the solution a obtained by (28), and $S^2(b)$ for beta-coefficients b (4). Using norms of vectors, we have

$$\begin{aligned} S^2(a) &= \|y - Xa\|^2 = \|(y - Xb) - X(a - b)\|^2 = \|y - Xb\|^2 - 2(a - b)'X'(y - Xb) \\ &\quad + (a - b)'X'X(a - b) = S^2(b) + (a - b)'C(a - b) \end{aligned} \tag{29}$$

In derivation (29) we use a property $X'(y - Xb) = 0$ satisfied due to solution (3). From (29) we get a relative index

$$S^2(a)/S^2(b) = 1 + ((a - b)'C(a - b))/(1 - R^2) \tag{30}$$

This expression defines efficiency of an adjusted regression versus least-squares model: if value (30) equals $(1 + d)$ per cent, than adjusted regression has d per cent bigger residual variance than regular regression. This is the price of the trade-off for an adjusted regression with interpretable coefficients and positive net effects.

5. NUMERICAL ESTIMATIONS

Formula (20) presents Shapley Value as a marginal input from each variable averaged by all possible coalitions. The important feature of this formula is the presentation of subsequent inputs of coalitions of the first, second, etc., levels to the total Shapley Value. If the data is available only on the several initial stages of coalitions with one, two, and some other subsets of variables, it is possible to use (20) for estimation of partial inputs to the total SV. Comparison of such cumulative values for each variable A, B, C, etc., allows to evaluate stability of the SV from partial data. This suggests an approach for reducing the computation time of the SV by limiting computation to the number of levels where stability is achieved. We can see by (20) that each term is constructed by calculating a mean value of combinations with the product and a mean value of combinations without it, thus, we can estimate those means by sampling combinations. This could be easily done and incorporated in the code whenever the number of regressors being evaluated is above 10 (see Reference [19]).

Let us consider an example of a real project on customer satisfaction study for a telephone customer service center. The variables are measured in scale from 1 (totally dissatisfied, or disagreed) to 7 (totally satisfied, or agreed). They include: y—overall satisfaction of clients with the company in general (dependent variable); $x1$—customer satisfaction with service representatives; $x2$—service representatives are courteous; $x3$—they provide all the needed information; $x4$—they give quick response; $x5$—they show care with customer problems; $x6$—they are accurate in the answers; $x7$—they take all the necessary actions. The data gathered by 242 respondents is available. All variables are positively correlated (pair correlations are from 0.52 to 0.89). The aim of the modeling was to measure the input of the predictors in their influence on the dependent variable.

In Table I some numerical results are presented. In the first row of this table we see that dependent variable is correlated with all the regressors rather evenly, so each variable can be more or less equally important in the model. However, by beta-coefficients and their t-statistics in the next two rows it is clear that most of the predictors are insignificant in the model. Two x's have negative beta-coefficients, and their net effects are negative too (next two rows in Table I), although pair correlations (and pairwise regressions of y by each x separately) have positive signs and similar values. Of course, it is the effects of multicollinearity, but knowing it does not help much in comparison of the variables' importance. And what do we do if we want to construct a pie-chart by shares of net effects with two of those negative? Multiple determination equals 0.356 and its t-statistics is $t_R = 11.52$ (see Table I), so the model is good for forecasting, but it is hardly useful for the analysis of the regressors.

Table I. Regressor coefficients and net effects.

Predictor	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	R^2
Correlation r_j	0.543	0.450	0.545	0.433	0.511	0.546	0.505	—
Beta (4)	0.255	– 0.022	0.177	– 0.039	0.052	0.195	0.029	0.356
t-statistics	2.26	– 0.25	1.28	– 0.45	0.50	1.49	0.28	11.52
Net effect (7)	0.139	– 0.010	0.096	– 0.016	0.027	0.106	0.014	0.356
Share, %	38.91	– 2.72	27.08	– 4.72	7.51	29.90	4.04	100
SV (21)	0.068	0.034	0.064	0.031	0.048	0.065	0.046	0.356
Share, % (22)	19.10	9.55	17.97	8.71	13.47	18.28	12.92	100
a_j (26)	0.125	0.075	0.117	0.072	0.094	0.119	0.092	0.356
a_j (28)	0.118	0.075	0.111	0.073	0.092	0.113	0.090	0.345

The next rows of Table I show the Shapley Value net effects—they are positive, thus, interpretable. Moreover, the net effects become closer to one another—it makes more sense taking into account similar values of the correlations with the dependent variable (Table I, the first row). These net effects can be used without any difficulties, particularly in graphical presentation of the regressors contribution into the model. Taking confidential probability $\alpha = 5$ per cent, we find by (23) confidential probability for net effects $\gamma = 0.73$ per cent, so $t_{\gamma/2} = 2.68$. The relative index δ^2 (25) equals 5.84 per cent, so all Shapley Value net effect shares are above this level (see Table I), thus, the contribution from each regressor is significant in the model.

Using Shapley Value, we also construct the adjusted coefficients of regression in the approach (26) and (28). We see (in the last two rows of Table I) that they are very close (although a_j (28) was obtained after several dozens of iterations by the 'nlminb' function in SPLUS software). Using the final set of coefficients we estimate that multiple determination for this model equals 0.345 (last row and last column in Table I)—it is just a little less than the original value of 0.356. Efficiency index (30) equals 1.017, i.e. residual variance of the adjusted regression is just 1.7 per cent higher than that of a regular regression, but at the same time all re-estimated parameters of regression become positive and interpretable.

Additionally, we use bootstrapping for evaluation of the considered characteristics—see Table II.

Again Shapley Value, based on the averaged sets of regressors, demonstrates stable results and robust decision rule. In Table II we represent beta-coefficients, net effects, Shapley Value net effects, and adjusted coefficients of regression (the first row in each group). By bootstrapping with 50 replications we estimated mean values and standard deviation of all characteristics—see the second and the third rows in each group of the results. The t-ratio of means to standard deviations is shown in the last row of each group. By the results in Table II we see a high volatility of beta-coefficients and regressor net effects—these characteristics are biased from their means, and the means are usually less than their standard deviations. At the same time all characteristics obtained via Shapley Value (SV net effects and adjusted coefficients of regression) are very stable. They are almost unbiased from their mean values, and these means are several times more than the corresponding standard errors. Therefore, when using SV net effects and adjusted coefficients of regression we can be sure of the contribution of individual regressors and in the estimation of the regressors' influence on the dependent variable.

Table II. Bootstrapping for regression coefficients and net effects.

Predictor	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$
Beta (4)	0.255	– 0.022	0.177	– 0.039	0.052	0.195	0.029
Mean	0.183	0.013	0.203	– 0.035	0.090	0.172	0.045
Std	0.246	0.170	0.210	0.127	0.178	0.187	0.154
t	0.74	0.08	0.97	– 0.27	0.51	0.92	0.29
NEF (7)	0.139	– 0.010	0.096	– 0.016	0.027	0.106	0.014
Mean	0.145	0.018	0.106	– 0.002	0.039	0.097	0.013
Std	0.135	0.071	0.121	0.061	0.139	0.108	0.084
t	1.07	0.25	0.88	– 0.03	0.28	0.90	0.15
SV (21)	0.068	0.034	0.064	0.031	0.048	0.065	0.046
Mean	0.071	0.040	0.071	0.035	0.057	0.74	0.053
Std	0.025	0.016	0.022	0.008	0.022	0.022	0.020
t	2.84	2.50	3.23	4.38	2.59	3.36	2.65
a_j (28)	0.118	0.075	0.111	0.073	0.092	0.113	0.090
Mean	0.112	0.072	0.107	0.074	0.099	0.113	0.086
Std	0.035	0.023	0.025	0.021	0.033	0.026	0.026
t	3.20	3.13	4.28	3.52	3.00	4.35	3.31

6. SUMMARY

We considered application of a tool from the cooperative game theory, namely, Shapley Value analysis, for evaluation of coefficients of regression and relative usefulness of the predictors in the model. The results are very encouraging—they show that it is possible to perform a reliable analysis even with a high degree of multicollinearity. The results can be understood by the specific structure of Shapley Value inputs as averages of the net effects over all possible coalitions of regressors. While regular regression coefficients and shares of importance are highly prone to multicollinearity distortions, all Shapley Value characteristics, being averaged values, are very consistent and demonstrate very stable bootstrapping output.

Due to several years of our experience actively using the described approach, the Shapley Value technique can be successfully combined with multiple regression, significantly facilitating analysis of the regression models in numerous practical applications.

ACKNOWLEDGEMENTS

The authors wish to thank a referee for the suggestions that improved the paper.

REFERENCES

1. Mason ChH, Perreault Jr WD. Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research* 1991; **28**:268–280.
2. Ferber R. *Marketing Research*. Ronald Press: New York, 1964.
3. Tishler A, Lipovetsky S. The flexible CES-GBC family of cost functions: derivation and application. *The Review of Economics and Statistics* 1997; **LXXIX**:638–646.

4. Tishler A, Lipovetsky S. A globally concave, monotone and flexible cost function: derivation and application. *Applied Stochastic Models in Business and Industry* 2000; **16**:279–296.
5. Kmenta J. *Elements of Econometrics*. Macmillan: New York, 1986.
6. Long JS. *Regression Models for Categorical and Limited Dependent Variables*. SAGE Publications: London, 1997.
7. Madigan D, Raftery AE. Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of American Statistical Association* 1994; **89**:1535–1546.
8. Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression model. *Journal of American Statistical Association* 1997; **92**:179–191.
9. Lipovetsky S, Conklin M. CRI: a collinearity resistant implement for analysis of regression problems. *Proceedings of the 31st Symposium on the Interface*: *Computing Science and Statistics*, Schaumburg, IL, 9–12 June, 1999; 282–287.
10. Lipovetsky S, Conklin M. Multiobjective regression modifications for collinearity. *Computers and Operations Research*, 2001, forthcoming.
11. Breiman L. Bagging predictors. *Machine Learning* 1994; **26**:123–140.
12. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 1995; **55**:119–139.
13. Lee SS, Elder JF. Bundling heterogeneous classifiers with advisor perceptrons. Elder Research Technical Report, University of Idaho, 1997.
14. Myerson RB. *Game Theory*: *Analysis of Conflict*. Harvard University Press: Cambridge, MA and London, England, 1991.
15. Owen G. *Game Theory*. Academic Press: New York, 1982.
16. Roth AE. The Shapley Value as a Von Neumann–Morganstern utility. *Econometrica* 1977; **45**:657–664.
17. Roth AE. (ed.). *The Shapley Value*: *Essays in Honor of Lloyd S. Shapley*. Cambridge University Press: Cambridge, 1988.
18. Conklin M, Lipovetsky S. Choosing product line variants: a game theory approach. *Proceedings of the 30th Symposium on the Interface*: *Computing Sciences and Statistics*: *Dimension Reduction*, *Computational Complexity and Information*. Minneapolis, MN, Vol. 30. 1998; 164–168.
19. Conklin M, Lipovetsky S. Modern marketing research combinatorial computations: Shapley Value versus TURF tools. *Proceedings of 1998 International S-Plus User Conference*, MathSoft Inc., Washington, DC, 8–9 October 1998.
20. Conklin M, Lipovetsky S. A winning tool for CPG″. *Marketing Research*: *A Magazine of Management and Applications* 2000; **11**:23–27.
21. Conklin M, Lipovetsky S. A new approach to choosing flavors. *Proceedings of the 11th Annual Advanced Research Techniques Forum of the American Marketing Association*, Monterey, CA, 4–7 June 2000.
22. Conklin M, Lipovetsky S. Identification of key dissatisfiers in customer satisfaction research. *The 11th Annual Advanced Research Techniques Forum of the American Marketing Association*, Monterey, CA, 4–7 June 2000.
23. Conklin M, Lipovetsky S. Evaluating the importance of predictors in the presence of multicollinearity. *Proceedings of the 12th Annual Advanced Research Techniques Forum of the American Marketing Association*, Amelia Island, FL, 24–27 June 2001.
24. Lipovetsky S, Tishler A. Linear methods in multimode data analysis for decision making. *Computers and Operations Research* 1994; **21**:169–183.
25. Tishler A, Lipovetsky S. Canonical correlation analysis for three data sets: a unified framework with application to management. *Computers and Operations Research* 1996; **23**:667–679.
26. Tishler A, Lipovetsky S. Modelling and forecasting with robust canonical analysis: method and application. *Computers and Operations Research* 2000; **27**:218–232.
27. Goldberger A. *Econometrics*. Wiley: New York, 1964.
28. Timm N. *Multivariate Analysis with Applications in Education and Psychology*. Brooks-Cole: Monterey, CA, 1975.
29. Darlington R. Multiple regression in psychological research and practice. *Psychological Bulletin* 1968; **79**:161–182.
30. Harris R. *A Primer of Multivariate Statistics*. Academic Press: New York and London, 1975.
31. Weisberg S. *Applied Linear Regression*. Wiley: New York, 1985.
32. Marquardt D. Generalized inverses, ridge regression and biased linear estimation. *Technometrics* 1970; **12**:591–612.
33. Hsu JC. *Multiple Comparisons*: *Theory and Methods*. Chapman & Hall: London, 1996.
34. Ehrenberg ASC. How good is best. *Journal of Royal Statistical Society*, *A*, 1982; **145**:364–366.
35. Cooley W, Lohnes P. *Multivariate Data Analysis*. Wiley: New York, 1971.
36. Grapentine T. Managing multicollinearity. *Marketing Research* 1997; 11–21.
37. Green PE, Carroll JD, DeSarbo WS. A new measure of predictor variable importance in multiple regression. *Journal of Marketing Research* 1978; **20**:356–360.