Estimators of Relative Importance in Linear Regression Based on Variance Decomposition

Ulrike GRÖMPING

Assigning shares of "relative importance" to each of a set of regressors is one of the key goals of researchers applying linear regression, particularly in sciences that work with observational data. Although the topic is quite old, advances in computational capabilities have led to increased applications of computer-intensive methods like averaging over orderings that enable a reasonable decomposition of the model variance. This article serves two purposes: to reconcile the large and somewhat fragmented body of recent literature on relative importance and to investigate the theoretical and empirical properties of the key competitors for decomposition of model variance.

KEY WORDS: Averaging over orderings; Linear model; Proportional marginal variance decomposition (PMVD); Sequential sums of squares.

1. INTRODUCTION

In many linear regression applications, a main goal of analysis is the determination of a ranking of the regressors or an explicit quantification of the relative importance of each regressor for the response. This type of application is often encountered in disciplines that rely on observational studies such as psychology, biology, ecology, economy and so forth (see, e.g., application areas in the references). If all regressors are uncorrelated, there is a simple and unique answer to the relative importance question. However, it is the very nature of observational data that regressors are typically correlated. In this case, assignment of relative importance becomes a challenging task, for which the standard output from linear regression models is not particularly well suited. This article focuses on relative importance assessment based on variance decomposition for linear regression with random regressor variables. Thus, metrics such as level importance [the product of the unstandardized regression coefficient with the regressor's mean, advocated, e.g., by Achen (1982, p. 71 ff)], which focus on the expected value of the response, are not covered here.

For reviews of work on relative importance, see Johnson and Lebreton (2004), Fickel (2001), Firth (1998), or Kruskal and Majors (1989). Historically, Darlington (1968) gave an insightful overview of the approaches to relative importance available at that time, which include, among others, the so-called "usefulness" of a regressor [which is the increase in R^2 if the regressor is added to the model that already includes all other regressors, equivalent to t statistics or so-called Type III sums of squares; see, e.g., Dobson (2002, p.88)], simple squared marginal correlations, squared standardized coefficients and products of standardized coefficients with marginal correlations (Hoffman 1960). All these approaches lead to the same result in case of uncorrelated regressors, but they can lead to quite different results for correlated regressors, and all these approaches have been criticized. Customers of statistical analysis often request a decomposition of the full model R^2 into contributions from the different regressors. Hoffman's (1960) proposal does deliver such decomposition, but is not considered appropriate by most authors mainly because some allocated contributions can become negative. Pratt (1987) provided a set of criteria under which Hoffman's proposal appears justified, which has convinced some authors to reconsider the method, while the present author, in line also with Darlington (1968), Bring (1996), or Johnson and Lebreton (2004), rejects this method as inappropriate (see Section 2.2).

The first proposal that will be pursued in this article came from Lindeman, Merenda, and Gold [1980, p. 119ff; made known by Kruskal (1987a,b), who mainly suggested a slightly different nonadditive variant], henceforth LMG, who proposed to use sequential [also called Type I, see, e.g., Dobson (2002, p. 88)] sums of squares from the linear model—the size of which depends on the order of the regressors in the model—and obtained an overall assessment by averaging over all orderings of regressors. This proposal has so far not found its way into the statistical mainstream, presumably for two reasons: first, it is computationally challenging and has become feasible only with enhanced computing power; second, its justification has originally been quite ad hoc, and its properties are not yet well understood. Nevertheless, there is a substantial amount of literature that proposes usage of this method under various names: Theil and Chung (1988) adopted the principle and proposed applying it to information rather than proportions of variation. In a frequently cited article, Chevan and Sutherland (1991) generalized the principle—

Ulrike Grömping is Professor for Business Mathematics and Statistics at Department II–Mathematics, Physics, and Chemistry, TFH Berlin, University of Applied Sciences, Luxemburger Strasse 10, 13353 Berlin, Germany (E-mail: *groemping@tfh-berlin.de*). The author thanks Norman Fickel, Barry Feldman, Ursula Garczarek, and Sabine Landau for useful discussions and comments. The constructive comments of both referees and the associate editor have also greatly improved the article.

 DOI: 10.1198/000313007X188252

called hierarchical partitioning in their paper—to more general classes of regression models, and Walsh and MacNally (2005) provided the R package `hier.part` for applying this approach. Budescu (1993) and Azen and Budescu (2003) introduced dominance analysis that—among other things—assigns exactly the LMG contribution to each regressor, and Azen (2003) provided a SAS macro for the relevant calculations. Lipovetsky and Conklin (2001) reinvented LMG from a game-theoretic perspective by applying the Shapley value (Shapley 1953). Stufken (1992) already noted that Chevan and Sutherland's proposal is equivalent to the Shapley value, and proposed that game theory might render useful additions to relative importance investigation.

The second proposal that will be discussed came from Feldman (2005) who introduced the "proportional marginal variance decomposition," henceforth PMVD, which is a weighted analogue of LMG with data-dependent weights. Feldman showed that PMVD is an instance of the game-theoretic proportional value (Feldman 1999; Ortmann 2000) and exploited this fact for simplifying computations. Grömping (2006) presented the R package `relaimpo` that calculates most of the metrics discussed so far, including PMVD.

There is a need in the applied sciences for appropriate methods of assigning relative importance, as evidenced by a recent boom of literature on relative importance (e.g., Soofi, Retzer, and Yasai-Ardekani 2000; MacNally 2000; Whittaker, Fouladi, and Williams 2002; Lebreton, Ployhart, and Ladd 2004; Johnson 2004; Budescu and Azen 2004; Conklin, Powaga, and Lipovetsky 2004; in addition to the already mentioned sources). Because of the lack of an accepted mainstream methodology for the important task of relative importance investigations, the field has substantially disintegrated, and many application areas appear to have reinvented the wheel in one way or other.

This article serves two purposes: (i) to reunite the relative-importance related aspects of the literature from the various fields and (ii) to investigate the statistical properties of the key competitors that decompose the model R^2 (LMG and PMVD). An introductory section on the linear model and on desirability criteria for relative importance metrics (Section 2) will be followed by an investigation of the theoretical quantities that are consistently estimated by LMG and PMVD (Section 3). Section 4 presents a simulation study on the distribution of the estimators, and Section 5 discusses the implications of the findings and areas of further research.

2. THE FRAMEWORK FOR VARIANCE DECOMPOSITION IN LINEAR REGRESSION

2.1 The Marginal Perspective on the Linear Regression Model

The focus in this article is on decomposing the variance of the response Y into proportions due to the X's (and error) in the linear regression model

$$Y = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_p + \varepsilon, \qquad \beta_0, \beta_1, \ldots, \beta_p \text{ fixed and unknown}, \quad (1)$$

where the random variables X_j, $j = 1, \ldots, p$, denote p regressor variables and the random variable ε denotes an error term with expectation 0 and variance $\sigma^2 > 0$ which is uncorrelated to the regressors. Because we assume a regression model with intercept, it can be assumed w.l.o.g. that all X's are centered (i.e., have expectation 0). The regressor variances are denoted as v_j, $j = 1, \ldots, p$, the inter-regressor correlations as ρ_{jk}, and the $p \times p$ covariance matrix between regressors is assumed to be positive definite so that any sample regressor matrix with $n > p$ rows is of full column rank with probability one. Model (1) implies the conditional moments $E(Y|\, X_1, \ldots, X_p) = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_p$ and $\text{var}(Y|\, X_1, \ldots, X_p) = \text{var}(\varepsilon|\, X_1, \ldots, X_p) = \sigma^2$ and the marginal variance model

$$\text{var}(Y) = \sum_{j=1}^{p} \beta_j^2 v_j + 2\sum_{j=1}^{p-1}\sum_{k=j+1}^{p} \beta_j\beta_k\sqrt{v_j v_k}\rho_{jk} + \sigma^2. \quad (2)$$

Throughout the article, the true coefficients $\beta_1, \ldots, \beta_p$ are those against which estimates from model (1) are consistent. If (1) is misspecified by omitting relevant variables, the true coefficients in this sense include the bias. Note that (2) depends on β_j and v_j through $\beta_j\sqrt{v_j}$ only, which is the coefficient one would obtain for the standardized regressor $X_j/\sqrt{v_j}$.

The first two summands of (2) constitute the part of the variance that is explained by the regressors, while the last summand is the error variance. R^2 from a linear model with n independent observations is consistent for the proportion of the first two summands of (2) in the total $\text{var}(Y)$. As long as the X's are uncorrelated with each other, the explained variance obviously decomposes into the contributions $\beta_j^2 v_j$, which can be consistently estimated using the unique sums of squares (SS) for each regressor.

In case of correlated X's, it is no longer obvious how (2) should be decomposed. LMG and PMVD choose different ways, which are detailed below. All discussions are in terms of the theoretical quantities towards which the estimates converge for increasing sample sizes, and w.l.o.g. allocation to the regressor X_1 is singled out for investigation. The order in which regressors are entered into the model is denoted as $r = (r_1, \ldots, r_p)$, which is a permutation of the regressors' indices $\{1, \ldots, p\}$, and the set of regressors appearing before X_1 in the order r is denoted as $S_1(r)$. In order to simplify formulas, let us further introduce the notations

$$\text{evar}(S) = \text{var}(Y) - \text{var}(Y|X_j,\, j \in S), \quad (3)$$

and

$$\text{svar}(M\,|S) = \text{evar}(M \cup S) - \text{evar}(S) \quad (4)$$

for the explained variance based on regressors with indices from S and the sequentially added explained variance when adding the regressors with indices in M to a model that already contains the regressors with indices in S. Note that the true coefficient of determination $R^2(S)$ can be written as $\text{evar}(S)/\text{var}(Y)$. For most purposes, working with evar is equivalent to working with R^2.

2.2 Desirability Criteria for Decomposition of R^2

The following criteria for decomposition of the model R^2 are considered useful in the literature, though seldom listed explicitly:

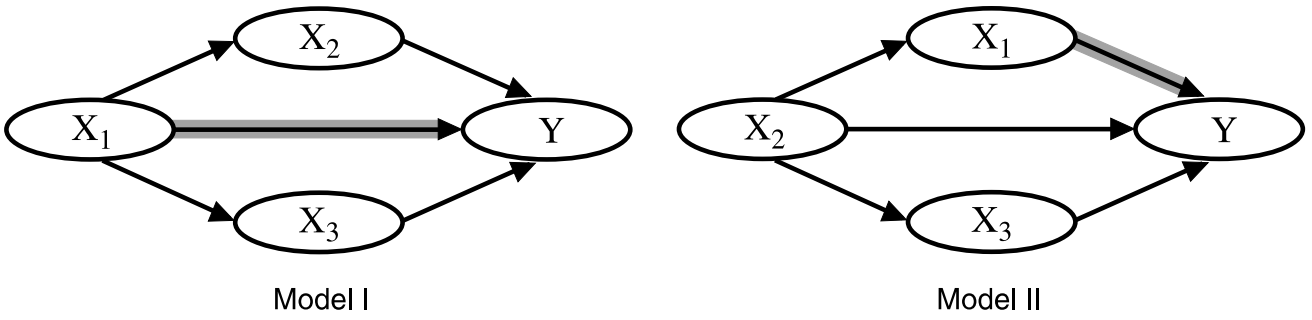

Figure 1. Two simple causal models that can lead to the linear regression model $E(Y|X_1, X_2, X_3) = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3$.

(a) *Proper decomposition*: the model variance is to be decomposed into shares, that is, the sum of all shares has to be the model variance.

(b) *Non-negativity*: all shares have to be non-negative.

(c) *Exclusion*: the share allocated to a regressor X_j with $\beta_j = 0$ should be 0.

(d) *Inclusion*: a regressor X_j with $\beta_j \neq 0$ should receive a nonzero share.

The list could be extended by further reasonable requests; the criteria listed here are the ones most relevant for comparing LMG and PMVD among each other and with other relative importance metrics. Criteria (a) to (d) are requested by various authors [see, e.g., Feldman 2005 (all), Darlington (1968, (a),(b)), Theil (1971, (b)), Johnson and Lebreton (2004, (a), (b)), Cox (1985, slightly different context, (c), (d))], and are cited here for ease of reference. Feldman (2005) postulated these four criteria in the sense of strict admissibility criteria and showed that PMVD is admissible in this sense while LMG is not. The present author agrees that criteria (a) and (b) are indispensable, so that the method by Hoffman (1960) as justified by Pratt (1987) is not further discussed because of its violation of criterion (b). Criterion (d) is fulfilled by both LMG and PMVD, and it is conjectured that it will be fulfilled by any nontrivial metric that fulfills both (a) and (b).

Exclusion (criterion (c)) is fulfilled for all relative importance metrics mentioned in this article—even the simple ones that have been severely criticized in the literature—as long as all regressors are uncorrelated. Feldman (2005) generally requested exclusion since he considered a regressor with a zero coefficient to be "spurious." When thinking of predictive relevance, a regressor with 0 coefficient in the equation does indeed not contribute anything useful, given that all regressors with nonzero coefficients are available, so that exclusion is a reasonable request. If the relative importance question is asked with a causal interpretation in mind (as is, e.g., the case when aiming at prioritizing intervention options) and regressors are correlated, exclusion is a less convincing requirement: Figure 1 shows graphs of two causal models—with directed arrows indicating a direct causal relation—that both (assuming linearity of all relations) imply the same linear regression model with $p = 3$ correlated regressors. In model I, regressor X_1 directly influences both other regressors and the response. If the shaded arrow is deleted from the graph, the coefficient β_1 becomes zero, since in the presence of X_2 and X_3 there is no additional explanatory value in X_1. Nevertheless, X_1 obviously exerts an influence on Y via the other two regressors, and there is no reason to request that it should be allocated a share of zero. In model II, X_2 and X_1 have swapped roles. Again, if the shaded arrow is deleted from the graph, the coefficient β_1 becomes zero. Now, it appears far more reasonable that X_1 should be allocated a share of zero.

As the linear regression model (1) is generally compatible with many different causal models, among them also those models for which exclusion is clearly unreasonable, exclusion does not appear to be a reasonable requirement for relative importance considerations, if causality considerations motivate the analysis.

Criteria (a) to (d) refer to properties of the theoretical quantities estimated and can also be applied for the estimated quantities, when replacing all theoretical values with their empirical counterparts. In addition to these criteria, a reasonably low variability of the estimators in cases of moderate multicollinearity is also an important aspect in assessing a method's performance.

3. WHAT DO RELATIVE IMPORTANCE METRICS ESTIMATE?

3.1 Estimated Quantities for LMG

As was mentioned before, LMG allocates to X_1 the average over allocations to X_1 from all possible orderings of regressors. In the simple case of two regressors, (2) simplifies to

$$\beta_1^2 v_1 + 2\beta_1\beta_2\sqrt{v_1 v_2}\rho_{12} + \beta_2^2 v_2 + \sigma^2. \qquad (2^*)$$

With n independent observations from the common distribution of Y, X_1, and X_2, let $\mathbf{y}$ denote the $n \times 1$-vector of centered responses, $\mathbf{x}_1$ the $n \times 1$-vector of centered values for regressor X_1, and the superscript T transposition and "-1" inversion. Then the model SS for X_1 in the role of the first and only regressor is

$$\mathbf{y}^{\mathrm{T}}\mathbf{x}_1\left(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1\right)^{-1}\mathbf{x}_1^{\mathrm{T}}\mathbf{y} = \left(\mathbf{x}_1^{\mathrm{T}}\mathbf{y}\right)^{\mathrm{T}}\left(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1\right)^{-1}\left(\mathbf{x}_1^{\mathrm{T}}\mathbf{y}\right),$$

which, when divided by n, is by simple considerations consistent for

$$\begin{aligned}\operatorname{svar}(\{1\}|\varnothing) &= \frac{\operatorname{cov}(Y, X_1)^2}{\operatorname{var}(X_1)} = \frac{\left(\beta_1 v_1 + \beta_2\sqrt{v_1 v_2}\rho_{12}\right)^2}{v_1} \\ &= \beta_1^2 v_1 + 2\beta_1\beta_2\sqrt{v_1 v_2}\rho_{12} + \beta_2^2 v_2\rho_{12}^2. \qquad (5)\end{aligned}$$

Note that—when alone in the model—the first regressor captures the full mixed term of the variance (2*) plus some of the unique

contribution of the second regressor in case of correlation. The contribution of the second regressor can be determined by the difference to the total model SS, which (divided by n) is consistent for $\text{var}(Y) - \sigma^2$. Thus, the variance attributed to X_2 after adjusting out X_1 becomes $\text{svar}(\{2\}|\{1\}) = \beta_2^2 v_2(1 - \rho_{12}^2)$. By analogy, $\text{svar}(\{1\}|\{2\}) = \beta_1^2 v_1(1 - \rho_{12}^2)$. Hence, the theoretical allocation of variance to X_1 estimated by LMG is the average over the two orders, that is,

$$\beta_1^2 v_1 + \beta_1\beta_2\sqrt{v_1 v_2}\rho_{12} + 0.5\left(\beta_2^2 v_2 - \beta_1^2 v_1\right)\rho_{12}^2. \quad (6)$$

Each regressor receives half of the mixed term in (2*). In addition, for $\rho_{12} \neq 0$ the regressor with larger $\beta_j^2 v_j$ donates part of its contribution to the regressor with smaller $\beta_j^2 v_j$. This third summand of (6) creates an equalization between correlated regressors with unequal $\beta_j^2 v_j$. In the light of the discussion of Figure 1, this can be seen as a precaution that takes care of the uncertainty regarding the underlying model structure. The third summand of (6) also causes LMG's violation of the exclusion criterion for correlated regressors: if $\beta_1 = 0$, $\beta_2 \neq 0$, $\rho_{12} \neq 0$, there will be a nonzero share allocated to X_1.

For p regressors, the LMG share allocated to X_1 is given as

$$\begin{aligned} \text{LMG}(1) &= \frac{1}{p!}\sum_{r\,\text{permutation}} \text{svar}(\{1\}|S_1(r)) \\ &= \frac{1}{p!}\sum_{S\subseteq\{2,\ldots,p\}} n(S)!(p-n(S)-1)!\text{svar}(\{1\}|S). \end{aligned} \quad (7)$$

All orders with the same $S_1(r)$ can be summarized into one summand. Thus, the computational burden is reduced from calculation of $p!$ summands to calculation of 2^{p-1} summands which are based on the 2^p quantities $\text{evar}(S)$ and $\text{evar}(S \cup \{1\})$, $S \subseteq \{2, \ldots, p\}$.

Some readers may find it more intuitive to think of LMG(1) as the average over model sizes i of average improvements in explained variance when adding regressor X_1 to a model of size i without X_1 (see Christensen 1992), that is,

$$\text{LMG}(1) = \frac{1}{p}\sum_{i=0}^{p-1}\left(\sum_{\substack{S\subseteq\{2,\ldots,p\}\\ n(S)=i}} \text{svar}(\{1\}|S)\Big/\binom{p-1}{i}\right). \quad (7^*)$$

We have already seen in the two-regressor case that LMG violates the exclusion criterion. The other three desirability criteria are satisfied, as can be easily verified from (7), noticing that the method averages non-negative contributions that sum to the total variance. Since (7) can be calculated for various scenarios, investigations into the behavior of the estimand are possible without simulation (see Section 3.4).

3.2 Estimated Quantities for PMVD

PMVD can be calculated as a *weighted* average of the same contributions averaged in (7). Each order of regressors receives a *data-dependent* weight. With weights $p(r)$ that will be discussed below, PMVD can be written as

$$\text{PMVD}(1) = \sum_{r\,\text{permutation}} p(r)\text{svar}(\{1\}|S_1(r)). \quad (8)$$

This formula is quite similar to (7), the difference lying in the weights $p(r)$, which preclude combining the summands with the same $S_1(r)$ into one. Note that (8) is inefficient for computation (for more efficient computation see Feldman 2005, Appendix A).

Definition of the weights $p(r)$ is as follows: If all regressors have nonzero coefficients, the permutation r receives a weight that is proportional to

$$\begin{aligned} L(r) &= \prod_{i=1}^{p-1}\text{svar}\left(\{r_{i+1},\ldots,r_p\}\,|\,\{r_1,\ldots,r_i\}\right)^{-1} \\ &= \prod_{i=1}^{p-1}\left(\text{evar}(\{1,\ldots,p\}) - \text{evar}(\{r_1,\ldots,r_i\})\right)^{-1}, \end{aligned} \quad (9)$$

that is, the weights are $p(r) = L(r)/\sum_r L(r)$, where summation in the denominator is over all possible permutations r. The factors in product (9) are increasing in size from $i = 1$ to $i = p - 1$. Weights are large, if the first regressor already captures a large portion of the explained variance (so that $(\text{evar}(\{1,\ldots,p\}) - \text{evar}(\{r_1\}))^{-1}$ is already relatively large). Also, if a set of regressors has a low explanatory value conditional on all other regressors, weights are large if all regressors from this set occur after the other regressors in the order. If some coefficients are zero, limiting considerations (see Feldman 2002) show that weights become positive for orderings with all 0-coefficient variables last, while any other ordering receives a weight of 0—in fact, the results for data with one or more coefficients estimated as 0 are identical to the results from models with the 0-coefficient variables omitted and their shares fixed at 0. Thus, PMVD weights guarantee exclusion, as they were designed to do. In addition, like any approach that can be written as an average over orderings, PMVD also guarantees the other three desirability criteria, using the same reasoning as for LMG.

For illustration of PMVD, let us apply (8) and (9) to a scenario with two regressors X_1 and X_2 and nonzero coefficients. The $L(r)$ consist of one factor only, with $L((1,2)) = \text{svar}(\{2\}|\{1\})^{-1}$, so that the weight $p((1,2))$ becomes

$$p((1,2)) = \frac{\text{svar}(\{1\}|\{2\})}{\text{svar}(\{1\}|\{2\}) + \text{svar}(\{2\}|\{1\})} = \frac{\beta_1^2 v_1}{\beta_1^2 v_1 + \beta_2^2 v_2}.$$

With $p((1,2))$ and $p((2,1))$ inserted in (8), using the sequential variances calculated in Section 3.1, the variance allocated to X_1 simplifies to

$$\beta_1^2 v_1 + \frac{\beta_1^2 v_1}{\beta_1^2 v_1 + \beta_2^2 v_2} * 2\beta_1\beta_2\sqrt{v_1 v_2}\rho_{12} \quad (10)$$

This result for two regressors has several specific properties none of which generalizes to $p > 2$: the share of the mixed term that a regressor receives is proportional to its individual term in the model. Also, the weight for order (1, 2) coincides with the proportion of R^2 allocated to X_1, and the weights do not depend on the correlation between the X's. For more than two regressors, the scenario investigations in Section 3.4 will shed further light on the behavior of PMVD.

Table 1. A few PMVD weight examples with $\text{corr}(X_1, X_2) = \text{corr}(X_2, X_3) = \rho$, $\text{corr}(X_1, X_3) = \rho^2$

	$\beta_1 = 1, \beta_2 = 1, \beta_3 = 1$			$\beta_1 = 5, \beta_2 = 4, \beta_3 = 3$			$\beta_1 = 4, \beta_2 = 1, \beta_3 = 0.3$		
Order r	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = -0.5$	$\rho = 0$	$\rho = 0.5$
(1,2,3)	0.174	0.167	0.129	0.344	0.320	0.257	0.855	0.859	0.828
(2,1,3)	0.109	0.167	0.210	0.154	0.235	0.296	0.043	0.058	0.073
(2,3,1)	0.109	0.167	0.210	0.056	0.085	0.107	0.000	0.000	0.000
(1,3,2)	0.217	0.167	0.161	0.242	0.180	0.181	0.096	0.077	0.093
(3,1,2)	0.217	0.167	0.161	0.135	0.110	0.105	0.005	0.005	0.005
(3,2,1)	0.174	0.167	0.129	0.069	0.070	0.054	0.000	0.000	0.000

3.3 Discussion of Weight Approaches

As mentioned before, LMG simply gives each order of regressors the same weight, that is, the weights are data-independent. If a predetermined order of variables can be specified, we have another situation of data-independent weights, for which one order has weight one, all others weight 0. Applied researchers facing a relative importance question often use an automated forward or backward selection of variables (e.g., Janz et al. 2001) to determine the order of regressors. This is an extreme case of data-dependent weights (1 for one order, 0 for all other orders), which guarantees exclusion in case of backward selection, but can deliver quite arbitrary allocations as is obvious from comparing $\text{svar}(\{1\}|\varnothing)$ to $\text{svar}(\{1\}|\{2\})$ for the two-regressor case (see (5) and subsequent calculations). Selection-based approaches have been criticized, for example, by Bring (1996) and references therein. The PMVD weights $p(r)$ can be seen as a compromise between this extreme form of data-dependent weights and equal weights for all regressors. They are concentrated on a few orders in case of very unequal β's and are more balanced between orders for constant or very similar β's. Table 1 shows a few examples for three regressors.

3.4 Scenario Investigations Regarding the Estimated Quantities

Trivially, LMG and PMVD coincide for uncorrelated regressors, since variance allocation is order-independent. Furthermore, symmetry considerations imply that both methods allocate equal shares for all regressors in case of equi-correlated regressors if all $\beta_j\sqrt{v_j}$ are identical. This is true for LMG because all orders generally receive the same weight. For PMVD, identical weights for all orders are guaranteed by Feldman's (2005) anonymity axiom in this situation. Combining these two situations, LMG and PMVD also coincide for several uncorrelated groups of equi-correlated regressors with group wise constant $\beta_j\sqrt{v_j}$. For any other situation, the two methods will yield more or less different results, as is exemplified in Figure 2.

Figure 2 depicts allocated shares according to LMG (thick lines) and PMVD (thin lines) for a group of four regressors for four different scenarios (a) to (d) ($v_j = \text{var}(X_j)$ fixed at 1). As was pointed out above, LMG and PMVD coincide for scenario (a) (identical β's with constant inter-regressor correlations). For scenario (b) (identical β's, $\text{corr}(X_j, X_k) = \rho^{|j-k|}$), both methods show a moderate dependence on the correlation parameter, where regressors X_2 and X_3 are ranked before X_1 and X_4 for positive ρ and vice versa for negative ρ. In the two scenarios with very unequal β's ((c) and (d)), LMG and PMVD behave quite different: For LMG, increasing absolute values of the correlation imply a substantial equalization of shares (note that the highest disparity is observed not for $\rho = 0$ but for small negative correlations). PMVD is far less sensitive to changes in ρ, and negative correlation parameters attenuate rather than reduce disparities between regressors.

Apparently, for scenarios with very similar β's or correlations close to 0, the difference between LMG and PMVD is small, while it becomes large for scenarios with large differences between β's or more extreme correlations. The ranking of regressors in terms of allocated shares is often but not always the same for both methods. In applications (positive correlations among X's, all β's same sign), the numerical differences between a few large and many smaller allocated shares have been observed to be more distinct for PMVD than for LMG; however, because of the higher variability of PMVD estimates (see next section), this has not paid off in terms of more statistically significant differences between allocated shares.

Note that the equalization of LMG shares for increasing positive correlation is caused by a continuity property of LMG: in the limit case for which all regressors are perfectly correlated, all regressors (reasonably) receive an equal share, since there is no information whatsoever on potential differences between regressors. LMG shares continuously approach this limit if the correlation matrix approaches a matrix of ones, which can be seen as a precaution taken because of increased model uncertainty (see also Section 2.2). PMVD does not possess a comparable continuity property; rather, the limit against which PMVD converges depends on the model coefficients.

4. SIMULATIONS

Following the discussion of the theoretical quantities for which the estimators are consistent, this section investigates the sampling distributions of the estimators, with particular focus on the dispersion of the estimators. For a multivariate normal distribution of $(Y, X_1, \ldots, X_p)^{\mathrm{T}}$, the asymptotic distribution of LMG is known [Budescu (1993) based on a result by Olkin and Siotani (1976)]. This result cannot be applied to PMVD, since the PMVD-weights are data-dependent. Also, it is desirable to not restrict results to multivariate normal regressors. Therefore, a simulation study has been conducted.

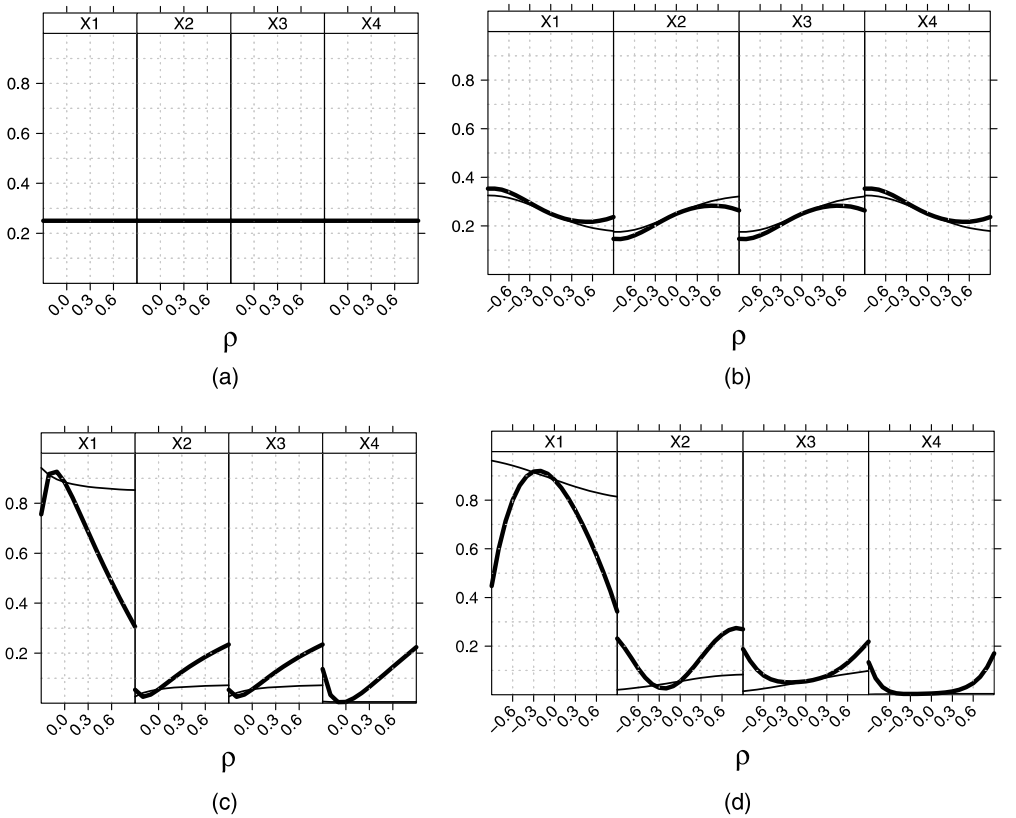

Figure 2. Proportions of R^2 allocated to each regressor for LMG (thick line) and PMVD (thin line).
Parameter vectors: (a) and (b): $\boldsymbol{\beta} = (1, 1, 1, 1)^{\mathrm{T}}$, (c) and (d): $\boldsymbol{\beta} = (4, 1, 1, 0.3)^{\mathrm{T}}$.
Correlations among $X_1, \ldots, X_4$: (a) and (c): $\mathrm{corr}(X_j, X_k) = \rho$, $j \neq k$. (b) and (d): $\mathrm{corr}(X_j, X_k) = \rho^{|j-k|}$.

Table 2. Simulation settings

Factor	Levels	
Correlation structure of $(X_1, \ldots, X_4)^{*)}$	∘ $\mathrm{corr}(X_j, X_k) = \rho$ for $j \neq k$	with $\rho = -0.3$ to 0.9 in steps of 0.1
	∘ $\mathrm{corr}(X_j, X_k) = \rho^{\lvert j-k \rvert}$	with $\rho = -0.9$ to 0.9 in steps of 0.1
Distribution of $(X_1, \ldots, X_4)^{**)}$	∘ Multivariate normal (expectations 0, variances 1)	
	∘ Exponential (= linear combination of exponentials; expectations depend on correlation structure, variances 1)	
Coefficient vectors $(\beta_1, \ldots, \beta_4)^{\mathrm{T}}$	∘ $\boldsymbol{\beta}_1 = (4, 1, 1, 0.3)^{\mathrm{T}}$	∘ $\boldsymbol{\beta}_5 = (1.2, 1, 1, 0.3)^{\mathrm{T}}$
	∘ $\boldsymbol{\beta}_2 = (1, 1, 1, 0.3)^{\mathrm{T}}$	∘ $\boldsymbol{\beta}_6 = (1, 1, 1, 0)^{\mathrm{T}}$
	∘ $\boldsymbol{\beta}_3 = (4, 1, 0, 0)^{\mathrm{T}}$	∘ $\boldsymbol{\beta}_7 = (4, 3.5, 3, 2.5)^{\mathrm{T}}$
	∘ $\boldsymbol{\beta}_4 = (1, 1, 1, 1)^{\mathrm{T}}$	
Sample sizes	∘ $n = 100$	independent observations
	∘ $n = 1000$	independent observations
True R^2 (controlled through σ^2)	∘ 0.25	
	∘ 0.5	
	∘ 0.9	

*) Observations on different units are independent. The correlation structure refers to the four regressors within each independently observed unit.

**) Note that the regressors in the exponential case have nonzero expectation, which is irrelevant, since it can be subsumed in an estimate of the intercept and does not affect any conclusions about explained variance.

4.1 The Simulation Set-ups

The simulations work with four regressors only, in order to allow various patterns and keep simulation times manageable. Each set-up has been replicated 500 times, which allows reasonably accurate estimates of the dispersion.

Table 2 shows the simulation settings. In the multivariate normal case, regressors have simply been simulated as independent multivariate normal samples with expectation vector $\mathbf{0}$ and covariance matrix as defined in the scenario, while the exponential case simulates the regressors as linear combinations of independent exponential random variables with expectation 1, such that the covariance matrix is equal to the correlation structure. In all simulations, the response has been simulated as $\beta_0 + X_1\beta_1 + \cdots + X_4\beta_4$ plus an independent normal error with expectation 0.

4.2 Distribution of the Estimated R^2 Percentages

Since the estimators are consistent for their respective true quantities, the average estimates are closer to the true values as discussed in Section 3 both with increasing sample size and increasing R^2. The simulations show no surprises here.

The most interesting outcome from the simulations is the assessment of the variability behavior of the allocated proportions of R^2. Figure 3 shows the interquartile ranges for LMG and PMVD for the most variable simulated scenario (lowest R^2, $n = 100$); although a discrete set of ρ values has been simulated, lines have been drawn for visual clarity. It can be seen that PMVD is most variable with positive correlations and sometimes with strong negative correlations, while LMG is most variable for small absolute values of correlations in most situations. PMVD is distinctly more variable than LMG for positive correlations, and sometimes also for strong negative correlations, while it is slightly less variable than LMG for moderate negative correlations in most cases. For $\boldsymbol{\beta}$-vectors with some zero elements ($\boldsymbol{\beta}_3$, $\boldsymbol{\beta}_6$), PMVD variability is very low for the respective shares (with a distinct advantage over LMG for scenarios with high R^2, not shown). This is in line with PMVD's property to satisfy the exclusion criterion. Also, PMVD shows lowest variability where true differences between coefficients are large with some coefficients relatively close to 0, while the variability disadvantage versus LMG is higher for $\boldsymbol{\beta}$-vectors with relatively similar coefficients. Overall, since variability differences in favor of PMVD are typically much smaller than those in favor of LMG, LMG is preferable in terms of variation.

5. DISCUSSION

This article has investigated two ways of decomposing R^2 in linear regression, LMG and PMVD. LMG has been reinvented numerous times by various researchers (see Section 1) and is based on the heuristic approach of averaging over all orders. Feldman (2005) criticized that LMG violates the exclusion criterion and designed PMVD specifically for satisfying the exclusion criterion, by employing a special set of data-dependent weights. While Feldman saw satisfaction of the exclusion criterion as so desirable that it was worth the price of increased computation efforts and increased variability of estimates, it has been pointed out in Section 2.2 of this article that exclusion is not a desirable criterion under all circumstances. If exclusion is considered an indispensable criterion for an application, PMVD must be used in spite of its larger variation and higher implementation effort. On the other hand, if a causal interpretation of the variance allocations is intended, LMG's equalizing behavior must be seen as a natural result of model uncertainty, and LMG is to be preferred. Luckily, in many (not all) applications the two methods

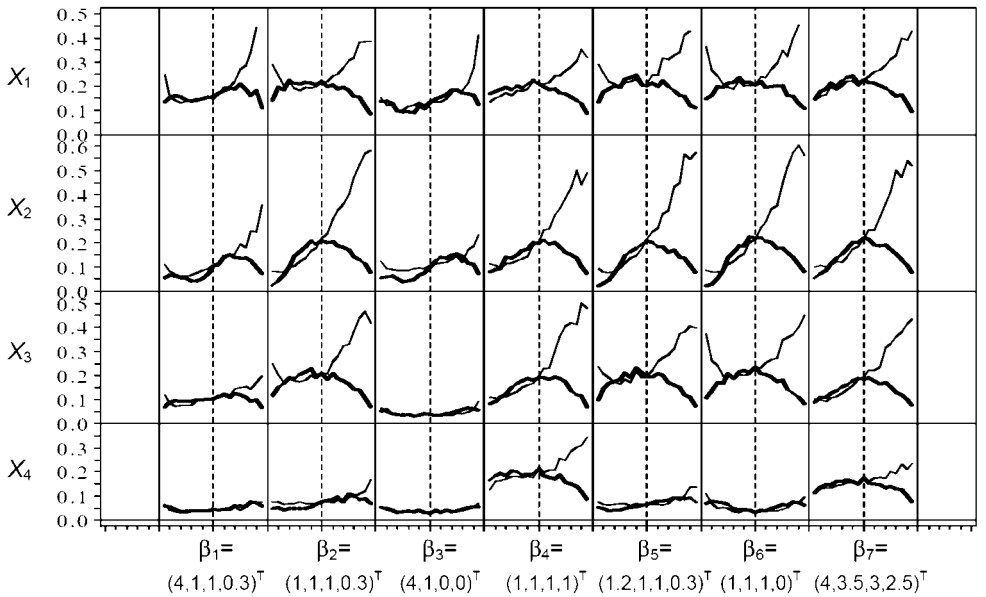

Figure 3. Interquartile ranges from 500 allocated proportions of R^2 for LMG (thick line) and PMVD (thin line).
Scenario: $R^2 = 0.25$, $n = 100$, normally distributed X's.
Correlations among $X_1, \ldots, X_4$: $\text{corr}(X_j, X_k) = \rho^{|j-k|}$ ($\rho = -0.9$ to 0.9 shown on horizontal axis for each $\boldsymbol{\beta}$, dashed vertical lines indicate $\rho = 0$).

give similar answers at least in terms of ranking regressors. It can be instructive to apply both methods and compare their results.

In many articles on relative importance, the methods are discussed in a purely descriptive fashion (also observed by Budescu and Azen 2004), that is, there is no discussion of true quantities estimated or statistical behavior of the variance contribution estimators. This is certainly at least partly due to a lack of understanding regarding the estimated quantities. This article provides a step towards improving this situation. Nevertheless, further investigations are needed in order to achieve a full understanding of variance decomposition in the sense of this article. Even though estimation is not the focus of most articles on relative importance, various researchers have recognized the need of reporting variability of assigned relative importances and have employed the bootstrap for determining confidence intervals (Azen's SAS macro, Grömping's R package). The bootstrap was also mentioned by Lipovetsky and Conklin (2001), Azen and Budescu (2003), and Johnson (2004), for example. Although not presented here, some simulation studies regarding coverage probabilities of bootstrap percentile confidence intervals for LMG have shown a somewhat anti-conservative behavior, with error levels up to about twice the nominal in some situations. Further investigations into the behavior of bootstrap confidence intervals are certainly needed.

Several authors—among them Ehrenberg (1990), Stufken (1992), and Christensen (1992)—have expressed reservations about the benefit of relative importance measures. Certainly, for a thorough understanding of any phenomenon, a detailed investigation of adequate data based on theory-driven explanatory models (e.g., path-analytical models in the spirit of Figure 1) is far more useful than a simplistic assessment of relative importance. In particular, a request for a decomposition of R^2 is often driven by a desire to prioritize intervention actions with the intention to influence the response. It is important to notice that any intervention bears the risk (or chance) of not only influencing the targeted regressor(s) but also the correlation structure among regressors. Thus, unexpected results may occur regarding changes of the response's variance. In this way, the benefit of the concept of decomposing R^2 is more limited than the typical user might realize. Nevertheless, if an agreed theory-driven explanatory model is unavailable, variables with high allocated shares of variability are natural candidates when trying to influence the response.

[Received February 2006. Revised December 2006.]

REFERENCES

Achen, C.H. (1982), *Interpreting and Using Regression*, Beverly Hills, CA: Sage.

Azen, R. (2003), "Dominance Analysis SAS Macros." Available online at *http://www.uwm.edu/~azen/damacro.html*.

Azen, R., and Budescu, D. V. (2003), "The Dominance Analysis Approach for Comparing Predictors in Multiple Regression," *Psychological Methods*, 8, 129–148.

Bring, J. (1996), "A Geometric Approach to Compare Variables in a Regression Model," *The American Statistician*, 50, 57–62.

Budescu, D. V. (1993), "Dominance Analysis: A New Approach to the Problem of Relative Importance in Multiple Regression," *Psychological Bulletin*, 114, 542–551.

Budescu, D. V., and Azen, R. (2004), "Beyond Global Measures of Relative Importance: Some Insights from Dominance Analysis," *Organizational Research Methods*, 7, 341–350.

Chevan, A., and Sutherland, M. (1991), "Hierarchical Partitioning," *The American Statistician*, 45, 90–96.

Christensen, R. (1992), Comment on "Hierarchical Partitioning," by Chevan and Sutherland, *The American Statistician*, 46, 74.

Conklin, M., Powaga, K., and Lipovetsky, S. (2004), "Customer Satisfaction Analysis: Identification of Key Drivers," *European Journal of Operational Research*, 154, 819–827.

Cox, L. A. (1985), "A New Measure of Attributable Risk for Public Health Applications," *Management Science*, 31, 800–813.

Darlington, R. B. (1968), "Multiple Regression in Psychological Research and Practice," *Psychological Bulletin*, 69, 161–182.

Dobson, A. (2002), *An Introduction to Generalized Linear Models* (2nd ed), London: Chapman and Hall.

Ehrenberg, A. S. C. (1990), "The Unimportance of Relative Importance," *The American Statistician*, 44, 260.

Feldman, B. (1999), "The Proportional Value of a Cooperative Game," unpublished manuscript for a contributed paper at the Econometric Society World Congress 2000. Available online at *http://netec.mcc.ac.uk*.

——— (2002), "A Dual Model of Cooperative Value," unpublished manuscript. Available online at *http://econwpa.wustl.edu:8089/eps/game/papers/0207/0207001.pdf*.

——— (2005), "Relative Importance and Value," unpublished manuscript (Version 1.1, March 19 2005). Available online at *http://www.prismanalytics.com/docs/RelativeImportance050319.pdf*.

Fickel, N. (2001), "Sequenzialregression: Eine neodeskriptive Lösung des Multikollinearitätsproblems mittels stufenweise bereinigter und synchronisierter Variablen," Habilitationsschrift, Berline: University of Erlangen-Nuremberg.

Firth, D. (1998), "Relative Importance of Explanatory Variables," Conference on Statistical Issues in the Social Sciences, Stockholm, October 1998. Online at *http://www.nuff.ox.ac.uk/sociology/alcd/relimp.pdf*.

Grömping, U. (2006), "Relative Importance for Linear Regression in R: The Package `relaimpo`," *Journal of Statistical Software*, 17, Issue 1. Available online at *http://www.jstatsoft.org/counter.php?id=167&url=v17/i01/v17i01.pdf&ct=1*.

Hoffman, P. J. (1960), "The Paramorphic Representation of Clinical Judgment," *Psychological Bulletin*, 57, 116–131.

Janz, K. F., Burns T. L., Torner J. C., Levy S. M., Paulos, R., Willing M. C., and Warren J. J. (2001), "Physical Activity and Bone Measures in Young Children: The Iowa Bone Development Study," *Pediatrics*, 107, 1387–1393.

Johnson, J. W. (2004), "Factors Affecting Relative Weights: The Influence of Sampling and Measurement Error," *Organizational Research Methods*, 7, 283–299.

Johnson, J. W., and Lebreton, J. M. (2004), "History and Use of Relative Importance Indices in Organizational Research," *Organizational Research Methods*, 7, 238–257.

Kruskal, W. (1987a), "Relative Importance by Averaging over Orderings," *The American Statistician*, 41, 6–10.

——— (1987b), Correction to "Relative Importance by Averaging Over Orderings," *The American Statistician*, 41, 341.

Kruskal, W., and Majors, R. (1989), "Concepts of Relative Importance in Recent Scientific Literature," *The American Statistician*, 43, 2–6.

Lebreton, J. M., Ployhart, R. E., and Ladd, R. T. (2004), "A Monte Carlo Comparison of Relative Importance Methodologies," *Organizational Research Methods*, 7, 258–282.

Lindeman, R. H., Merenda, P. F., and Gold, R. Z. (1980), *Introduction to Bivariate and Multivariate Analysis*, Glenview, IL: Scott, Foresman.

Lipovetsky, S., and Conklin, M. (2001), "Analysis of Regression in Game Theory Approach," *Applied Stochastic Models in Business and Industry*, 17, 319–330.

MacNally, R. (2000), "Regression and Model Building in Conservation Biology, Biogeography and Ecology: The Distinction Between and Reconciliation of 'Predictive' and 'Explanatory' Models," *Biodiversity and Conservation*, 9, 655–671.

Olkin, I., and Siotani, M. (1976), "Asymtotic Distribution of Functions of a Correlation Matrix," in *Essays in Probability and Statistics*, ed. S. Ideka, Tokyo: Shinho Tsusho, pp. 235–251.

Ortmann, K. M. (2000), "The Proportional Value of a Positive Cooperative Game," *Mathematical Methods of Operations Research*, 51, 235–248.

Pratt, J. W. (1987), "Dividing the Indivisible: Using Simple Symmetry to Partition Variance Explained," in *Proceedings of Second Tampere Conference in Statistics*, eds. T. Pukkila and S. Puntanen, University of Tampere, Finland, pp. 245–260.

Shapley, L. (1953), "A Value for n-Person Games," reprinted in *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, ed. A. Roth, Cambridge: Cambridge University Press.

Soofi, E. S., Retzer, J. J., and Yasai-Ardekani, M. (2000), "A Framework for Measuring the Importance of Variables with Applications to Management Research and Decision Models," *Decision Sciences*, 31, 1–31.

Stufken, J. (1992), "On Hierarchical Partitioning," *The American Statistician*, 46, 70–71.

Theil, H. (1971), *Principles of Econometrics*, New York: Wiley.

Theil, H., and Chung, C.-F. (1988), "Information-Theoretic Measures of Fit for Univariate and Multivariate Linear Regressions," *The American Statistician*, 42, 249–252.

Walsh, C., and MacNally, R. (2005), "`hier.part`: Hierarchical Partitioning," R-package version 1.0-1. Available online at *http://cran.r-project.org/doc/packages/hier.part.pdf*.

Whittaker, T. A., Fouladi, R.T., and Williams, N. J. (2002), "Determining Predictor Importance in Multiple Regression Under Varied Correlational And Distributional Conditions," *Journal of Modern Applied Statistical Methods*, 1, 354–366.