Relative Importance and Value[1]

Barry Feldman

Prism Analytics
53 West Jackson Blvd. #724
Chicago, IL 60604
barry@prismanalytics.com

Kellstadt Graduate School of Business
DePaul University
1 East Jackson Blvd. #6100
Chicago, IL 60604
bfeldma2@depaul.edu

MARCH 19, 2005
PRELIMINARY - PLEASE DO NOT QUOTE WITHOUT PERMISSION
QUESTIONS AND COMMENTS WELCOME
VERSION 1.1

ABSTRACT

Three axioms provide a formal definition of relative importance in a statistical or econometric model by identifying the likelihood that any ordering of independent variables is correctly ordered with respect to their relative importance. The expected contribution to model performance of independent variables with respect to this distribution is the proportional marginal decomposition of model performance with respect to the performance measure. Decomposition components are shown to be equal to the proportional value (Ortmann (2000), Feldman (1999, 2002)) of an appropriately constructed cooperative game. Also addressed are admissibility criteria for measures of relative importance, other measures of relative importance, examples, procedures for constructing confidence intervals, and extensions and limitations.

---

1 Introduction

This paper provides a precise and flexible definition of the relative importance of explanatory variables in statistical and econometric models. It also proposes four fundamental criteria that measures of relative importance should meet. The definition of relative importance is based on three axioms that identify a probability distribution over the possible orderings of variables in a model. The resulting method, proportional marginal decomposition (PMD), may be used with a wide variety of models and performance measures, including least squares and maximum likelihood.

Firth (1998) traces the concept of relative importance as far back as Hooker and Yule (1908). A need for better measures of statistical relative importance has recently been expressed in many fields. Healy (1990) and Schemper (1993) in medicine, Frees (1998) addressing insurance risk analysis, Soofi, Retzer and Yasai-Ardekani (2000) in management science and Kruskal and Majors (1989) in social science are examples. Many proposed measures have been criticized.[2] While general measures of relative importance have not found wide-spread acceptance, specialized methods have found increasing acceptance, for example, in time series analysis since Sims (1980).[3]

In practice, statistical significance measures such as t-statistics are widely used as *de facto* measures of relative importance. Elementary considerations dictate that statistical significance measures are not reliable measures of relative importance. For example, the joint marginal contribution to model performance of two explanatory variables increases with their mutual correlation. However, their marginal contributions to explained variance and, thus, their statistical significance levels decline. Relative importance is a measure of *full contribution.*

The most important direct purpose of a measure of statistical relative importance is to reduce the time, effort and skill required to identify and assess the effect of the joint correlations present among explanatory variables. A useful summary measure of relative importance is of obvious utility.

The principal roadblock to development of acceptable general measures of relative importance is that pure statistical theory has been an insufficient basis. But this should not foreclose the possibility that useful, if imperfect, methods might be developed through other means. Any map-maker's map necessarily distorts the curvature of the earth's surface, yet maps are ubiquitous. The admissibility criteria advanced in this paper provide a set of minimal standards for acceptable measures of relative importance.

The most frequently proposed measure of relative importance has been variance decomposition by averaging marginal contributions of independent variables over all

---

[2]King (1987) provides an iconoclastic criticism on the inappropriate use of standard regression coefficients, bivariate correlation and partial correlation. Both Heckman (1995) and Goldberger and Manski (1995) severely criticize the *ad hoc* use of relative importance measures. Ehrenberg's (1990) letter "The unimportance of relative importance," is disdainful of the concept of relative importance in statistics.

[3]See, for example, Campbell (1992).

orderings of variables. The *averaging method* appears to have been first proposed by Lindeman, Merenda and Gold (1980). Kruskal (1987) and Chevan and Sutherland (1991) propose essentially the same method. Soofi, Retzer and Yasai-Ardekani (2000) show that averaging is a maximum entropy estimator. Averaging is shown here to violate the *proper exclusion* criterion for measures of relative importance. A variable with a true beta of zero can have a positive relative importance.

The most widely used measure of relative importance today appears to be the well known variance decomposition $CVD_i = \beta_i \sum \beta_j \sigma_{ij}$. I refer to this method as *covariance decomposition*. Pratt (1987) provides an axiomatic characterization. Covariance decomposition is widely used commercially, for example, in current portfolio management and analysis software.[4] It violates two proposed relative importance criteria. Variables with non-zero true betas can be assigned zero relative importance. This violates the *proper inclusion* criterion. Covariance components can also be negative, violating the *nonnegativity* criterion.

The measures of relative importance considered hare have a relationship to cooperative game and bargaining theory. This relationship appears to have first been noted by Stufken (1992).[5] Define a *statistical cooperative game* by making an equivalence between the independent variables in a model and the players in a cooperative game. The coalitions in the game then represent all possible subsets of variables. If the worth of a coalition is defined as the marginal contribution to explained variance of its variables, then the Shapley (1953) value of the game is equal to the result of the averaging method. If, instead, the worth of a coalition is the model imputed aggregate variance for these variables (*i.e.*, $\beta'_S \Sigma_S \beta_S$, where the subscript represents the restriction of the estimated full model parameters to variables in S), then the Shapley value of the game is the model's covariance decomposition. Further, if the worth of a coalition is again defined by the marginal contribution to explained variance, then the proportional value (Ortmann (2000) and Feldman (1999, 2002)) of this game is equal to the model's proportional marginal variance decomposition.

In addition to being an interpretive aid, measures of relative importance may be of practical use in some aspects of model construction. For example, in data-based model decisions, simple rules incorporating relative importance measures are likely to result in more robust models.

Section 2 of this paper presents the proposed admissibility criteria for measures of relative importance. Section 3 presents the notational framework, axioms, and fundamental results defining proportional marginal decomposition. Section 4 derives properties of the measures of relative importance considered in this paper and, particularly, their admissibility according to the criteria of Section 2. Section 5 concerns application and presents several examples, including an extend example based on the analysis of a hedge fund. The empirical sample distribution properties of relative importance measures are also examined. Section 6 provides discussion about applications and extensions. Section 7 is the conclusion.

---

[4]Providers include Barra, Factset and Northfield.

[5]See Myerson (1991) for an introduction to cooperative game theory.

2 Admissibility Criteria

The standards for admissibility criteria must necessarily be considerably greater than for axioms. Axioms must withstand a test of reasonableness. It must be considered unreasonable that a measure of relative importance violates an admissibility criterion. These criteria withstand this test.

Let Θ be a statistical model with model performance measure μ, estimated parameter vector β and true but unobserved parameter vector β^*. Consider a measure of relative importance $\phi(\Theta, \mu, \beta)$.

1. **Nonnegativity.** All decomposition components must be greater or equal to zero: $\phi_i(\Theta, \mu, \beta) \geq 0$ for variable i in the model.

2. **Proper exclusion.** Spurious variables included in a model should receive no decomposition share. If $\beta_i^* = 0$ then $\phi_i(\Theta, \mu, \beta^*) = 0$.

3. **Proper inclusion.** Variables properly part of the model should receive a decomposition share. If $\beta_i^* \neq 0$ then $\phi_i(\Theta, \mu, \beta^*) \neq 0$.

4. **Full contribution.** Relative importance must measure total contribution to model performance. Let S be a set of variables such that for a variable i in S and any variable j not in S, $\sigma_{ij} = 0$. Then $\sum_{i \in S} \phi_i(\Theta, \mu, \beta^*) = w(S)$, where $w(S)$ is the joint marginal contribution to model performance of the variables in S. The sum of relative importance components of the variables in S must equal their joint marginal contribution to model performance.

Nonnegativity is perhaps the most basic criterion. Relative importance may be understood as a measure of the relative information contributed to the model by a variable. Information is inherently nonnegative.[6] Similarly, measures of model performance are inherently nonnegative as well.

Proper exclusion is necessary for consistency with statistical theory. What would it mean that a variable with a true beta of zero had positive relative importance? (Note, the definition concerns the true and not the sampled-based value of the relative importance measure.) Proper *inclusion* is also necessary for consistency with statistical theory. What would it mean that a variable with nonzero true beta had zero relative importance?

Full contribution is the essence of relative importance. It is the only criterion unavoidably violated by statistical significance measures. Full contribution also requires efficiency of a relative importance estimator in the game theoretic sense. The explanatory power of the complete model based on all variables must be divided among

---

[6] See Theil (1987) and Theil and Chung (1988) regarding an information theoretic interpretation of model performance measures. Note, also, Theil's (1971) comments regarding the unsuitability of covariance decomposition as a measure of relative importance due to violation of this criterion (p. 167, Section 4.2).

these. The conditions for application of full contribution are trivially satisfied in this case since there are no variables that are not in the set of all variables.

These criteria appear unobjectionable. It is difficult, also, to identify other qualities of a measure of relative importance that should also be considered essential. For example, some might wish to make linearity a criterion. Linearity may be unobjectionable as an axiom. However, there appears to be no compelling statistical or information theoretic basis to consider it a required property of an admissible estimator.

3   Proportional Marginal Decomposition

The basic framework is that of a probability distribution over the set of all possible permutations – or orders – of variables in a statistical model. The intuitive understanding of the probability associated with any ordering is that it indicates the likelihood that the ordering is a correct ordering of the variables in terms of their relative importance. The exact meaning of relative importance is determined by the axioms that identify the probability distribution.

3.1   Analytic Framework

Consider a model Θ with n independent variables $N = \{1, 2, \ldots, n\}$ and a single dependent variable y. Sets of independent variables are represented by $S \subseteq N$. Let μ be a measure of model performance such as log likelihood or explained variance (or R^2) of the model. Model performance conditional on restricting the independent variables to the set S will be indicated by $\mu_\Theta(S)$. Acceptable measures of model performance must be weakly monotonic, that is, if $S \subset T$, then $\mu_\Theta(S) \leq \mu_\Theta(T)$. Performance measures associated with model likelihood or statistical tests are natural model performance measures.

Marginal contributions to model performance are the basis of many statistical tests, and, particularly the likelihood ratio test and the F-test. Define the function w on all subsets of N as follows:

$$w(S) = \mu_\Theta(N) - \mu_\Theta(N \setminus S). \tag{1}$$

Note that $w(S) \geq 0$ for all S by construction and that w is a cooperative game. Then $w(S)$ is called the *worth* of the set S.

Let $\mathcal{R}(N)$ be the set of all $n!$ orderings of the variables of N and let $r = (r_1, r_2, \ldots, r_n) \in \mathcal{R}(N)$ be an ordering. Define S_k^r to be the set of the first k variables in the order r. If S has k variables, S is *included* in r if and only if $S = S_k^r$. That is, a set of k variables is included in an order r if and only if the first k variables in r are all in the set S. The notation $S \in r$ indicates S is included in r. Let $r(i)$ be the position of variable i in order r, so $i = r_{r(i)}$.

Relative importance may now be formally defined as a binary relation. If variable j is more important than variable i, write $i \prec j$. An ordering $r \in \mathcal{R}(N)$ will be considered to be in order of increasing relative importance if $r_1 \prec r_2 \prec \cdots r_{n-1} \prec r_n$.

Take an $r \in \mathcal{R}(N)$. Consider that r defines an order of entry of variables into the model Θ. Perhaps awkwardly, the order is the reverse of the variable ordering in r. First r_n enters, then r_{n-1}, and r_1 is the last to enter. This convention has the result that $w(S_i^r)$ is the marginal contribution of the last i variables to enter Θ in order r. We can also define the joint marginal contribution vector of included sets to be $MC(r) = (w(S_i^r))_{i=1}^n$, the vector of marginal contributions of the sets included in r.

The individual positional marginal contribution to performance of the variable in position i relative to order r is represented by $M_i(r)$, where

$$M_i(r) = w(S_i^r) - w(S_{i-1}^r), \tag{2}$$

$S_0^r = w(\emptyset) = 0$. The positional marginal contribution of a variable i is then represented by $M_{r(i)}(r)$. Marginal contributions are also defined relative to sets, so that the marginal contribution of variable i to $w(S)$ is $M_i(S) = w(S) - w(S \setminus i)$.

Let $\mathbf{p} \in \Delta_{n!}$ be the relative importance probability distribution to be defined on $\mathcal{R}(N)$. Rather than work with $\mathbf{p}$ directly, it will be easier to work with the likelihood function $L(r)$ implicitly defined as follows:

$$\mathbf{p}(r^*) = \frac{L(r^*)}{\displaystyle\sum_{r \in \mathcal{R}(N)} L(r)}, \tag{3}$$

where $L(r) \geq 0$.

The expected marginal contribution to model performance of a variable i with respect to $\mathbf{p}$ is then

$$\phi_i(w) = E_{\mathbf{p}}\left[M_i(r)\right] = \sum_{r \in \mathcal{R}(N)} \mathbf{p}(r) M_{r(i)}(r). \tag{4}$$

3.2 Axioms

Assume order r has the property that $M_1(r) < M_2(r) < \cdots < M_n(r)$, that is the positional marginal contributions to model performance of individual variables are strictly increasing. Then it might be reasonable to conclude that it is relatively likely that r is correctly ordered according to increasing relative importance and $r_1 \prec r_2 \prec \cdots \prec r_n$. The question, however, is how much more likely is it that r is correctly ordered than another ordering r^*? The first task is to define a reasonable set of axioms that allow us to answer this question.

The most obvious necessary property of any decomposition procedure should be that when all n variables are orthogonal, the shares ϕ_i for each variable i must be its

5

individual marginal contribution $M_i(\bar{i}) = w(\bar{i})$. But orthogonality implies $M_{r(i)}(r) = M_i(\bar{i}) = w(\bar{i})$ for all r. Thus every probability distributions on $\mathcal{R}(N)$ must give the correct expectation according to equation (4) when variables are orthogonal. Call this the *random order orthogonal decomposition property*. No axiom is needed.

3.2.1 Anonymity

The first axiom is essentially technical in nature. It requires that the likelihood $L(r)$ should depend only on sets included in r, and, further should depend only on $w(S)$ for $S \in r$ and not the identity of S or the variables included in S. These properties are reflected in the axiom of anonymity.

Axiom 3.1 ANONYMITY: *If $MC(r^*) = MC(r)$ then $L(r^*) = L(r)$.*

Clearly, $w(S)$ for $S \notin r$ influence $\mathbf{p}(r)$ through equation (3). It is thus redundant that these effects should also enter directly through $L(r)$. The requirement that only $w(S)$ and not the identity of S should influence $L(r)$ is similar to the better known symmetry axiom that is used in the identification of the Shapley value. The effect is to eliminate likelihood functions that vary with the identity of variables in particular positions in an ordering.

3.2.2 Proper Exclusion

The second axiom is a version of the proper exclusion criterion of Section 2. The limit condition formulation is a technical device necessary to avoid division by zero. The meaning and effect is the same as the simpler proper exclusion condition. A variable with a zero beta should receive zero decomposition share.

Axiom 3.2 LIMIT PROPER EXCLUSION: *Let w be defined by a model Θ and performance measure μ where $\beta_i^* = 0$. Consider a sequence of games w_k, $k = 1, 2, \ldots, \infty$, where w_k is based on $\beta_j^k = \beta_j^*$ for $j \neq i$, $\beta_i^1 > 0$, and $\beta_i^k \to 0$. Then*

$$\lim_{k \to \infty} \varphi_i(w_k) = 0.$$

3.2.3 Equal Proportional Effect

Finally, consider how $L(r)$ should change with respect to changes in $MC(r)$, the marginal contributions of sets included in r. I propose that a change in marginal contribution of any set of variables S included in r should result in a change in likelihood of equal proportional magnitude. Absolute values are considered because the sign the change is determined by limit proper exclusion.

Axiom 3.3 EQUAL PROPORTIONAL EFFECT: $\left| \dfrac{\partial \ln L(r)}{\partial \ln w(S)} \right| = 1.$

Equal proportional effect is a simple and natural approach to consistently take into account the marginal contributions to model performance of all included sets. The obvious alternative choice would to assume a linear relationship. The result would be to bias the measure toward the contributions of large sets with relatively larger marginal contributions. Attempts could be made to correct for the number of variables in a set. For example, the change in likelihood could be linear in the per-variable marginal contribution of a set. Such an approach does not fit well with the typically nonlinear decline in joint marginal contribution to explained variance as the number of variables declines. Note that the assumption of a non-zero linear effect could not identify the averaging approach as it requires that $MC(r)$ have no effect on $L(r)$.

The most questionable aspect of the axiom of equal proportional effect is not the choice of proportionality, but rather the requirement of equality of proportional effects. Why should the magnitude of the proportional change in likelihood *not* be twice the magnitude of the proportional change in marginal contribution? Or why should it be fixed at all and not estimated from Θ?

If the relationship is to be fixed, then a relationship of equality is natural if only because there is no apparent rationale for any other choice. Estimating a coefficient $\alpha = \partial \ln L(r)/\partial \ln w(S)$ from the data has an appeal to it, but Section 6.2 argues that allowing α to depend on the data is inappropriate because then relative importance would then effectively be a function of statistical significance.

3.3   The Main Result

Our first step in assembling the implications of these axioms is to observe that limit proper exclusion requires the $\partial \ln L(r)/\partial \ln w(S) < 0$. This can be seen by considering the marginal contribution of a variable i. Assume that i is first in order r: $S_1^r = i$. Assume the marginal contribution $w(S_1^r)$ becomes smaller. Then $\mathbf{p}(r)$ must grow relative to the probability of orders r^* where i is not first if the expected contribution of i is to decline.

Next, since $dX/dY = dX/d\ln Y(1/Y)$, limit proper exclusion and equal proportional effect together requires that $-\partial \ln L(r)/\partial w(S) = 1/w(S)$. This then implies

$$- \ln L(r) = \ c_r \ + \sum_{S \in r} \int_{x=0}^{w(S)} \frac{1}{x} \, dx = \ c_r \ + \sum_{S \in r} \ln w(S), \tag{5}$$

where c_r is a possible multiplicative factor consistent with limit proper exclusion and equal proportional effect axioms. Anonymity requires that c_r is constant for all $r \in \mathcal{R}$. Because c_r is a constant and $L(r)$ appears in both the numerator and denominator of equation (3), c_r may be assumed equal to zero.

Multiplying both sides of (5) by (-1) and taking antilogs leads to the following.

Lemma 3.1 *Given a model Θ, a performance measure μ and a game w generated by*

μ, the likelihood determined by the axioms of anonymity, limit proper exclusion and equal proportional effect is

$$L(r) = \left(\prod_{S \in r} w(S) \right)^{-1}.$$

Define the normalizing factor $P(N)$ as follows:

$$P(N) = \left(\sum_{r \in \mathcal{R}(N)} L(r) \right)^{-1} \tag{6}$$

The probability in the expectation (4) is then defined as $\mathbf{p}(r) = P(N)L(r)$ and substitution into (4) yields the following representation of the expectation defining proportional marginal decomposition:

$$\varphi_i(w) = P(N) \sum_{r \in \mathcal{R}(N)} L(r) M_{r(i)}(r). \tag{7}$$

The normalizing factor (6) for sets formed by removing one variable from N is easily proved to have a direct relationship to likelihoods and marginal contributions defined on $\mathcal{R}(N)$.

Lemma 3.2 *For any $i \in N$,*

$$P(N \setminus i) = \left(\sum_{r \in \mathcal{R}(N)} L(r) \, M_{r(i)}(r) \right)^{-1},$$

Proof: See Appendix B.

This result is of significance because the definition of $P(S)$ for any $S \subset N$ is the same as one representation of the ratio potential $P(S, w)$ of a cooperative game.[7] Substitution of the result of Lemma 3.2 into equation (7) yields the relationship $\varphi_i(w) = P(N)/P(N \setminus i)$. This is the definition of the proportional value (*e.g.*, Ortmann (2000) Definition 2.2, Feldman (2002) Equation (2.3)), and proves the following theorem.

---

[7]Feldman (2002) Lemma 2.1 derives this representation. See Appendix A for a description of the ratio potential. The concept of the potential of a cooperative game is introduced in Hart and Mas-Colell (1989), who define a linear potential that generates the Shapley value. Feldman (1999, 2002) and Ortmann (2000) develop the ratio potential that generates the proportional value.

Theorem 3.1 *The decomposition of the performance of a model Θ based on independent variables N, performance measure μ, cooperative game w defined by N and μ, and relative importance axioms anonymity, limit proper exclusion and equal proportional effect, $\varphi(w)$, is the proportional value of w.*

The representation of the proportional value of equation (7) is, in fact, equivalent to the random order representation provided by Lemma 2.9 of Feldman (2002). The proof of Lemma 3.2 here provides a direct and more transparent route to this result.

3.4   Computing the Proportional Value When a $w(S) = 0$

The case of computing the proportional value and a proportional marginal decomposition when the marginal contribution of at least one set S is equal to zero must be considered. Since $L(r) = \infty$ for any r that includes S with $w(S) = 0$, the probability distribution is not defined. This situation may be addressed by considering the game w_ϵ defined with respect to an $\epsilon > 0$ by the relation $w_\epsilon(S) = w(S) + |S|\,\epsilon$. To $w(S)$ is added ϵ times the number of variables in the set. Then $\varphi(w)$ is defined by the relation

$$\varphi(w) = \lim_{\epsilon \to 0} \varphi(w_\epsilon). \tag{8}$$

This result is based on Feldman (2002) Theorem 6.1, which provides sufficient conditions for this limit to be well defined. Basically, the limit is well defined except in the circumstance that there is an S with $|S| > 1$, with $w(S) > 0$, and $w(R) = 0$ for all proper subsets of S. In empirical work it is virtually impossible to find arrive at this condition except in cases of perfect multicolinearity among a set of independent variables.

3.5   Estimation Consistency

For any model Θ and performance measure μ the sample value of φ is defined by the expectation (4). In practice we must be concerned with the relationship between sample values of φ and its true value. Given the nonlinearity of φ it is clear that the expected value of φ is not, in general, its true value. This is obvious because the expectation of the product of two random variables is not the product of their expectation. It is however, easy to show that sample values of φ are consistent estimators of the true value.

Lemma 3.3 *Consider a model Θ with performance measure μ where the true parameter values β^* are consistently estimated by β. Then $\operatorname{plim} \varphi(w) = \varphi^*(w)$.*

Proof: Since $\operatorname{plim} \beta = \beta^*$, Slutsky's theorem on the consistency of functions of consistent random variables leads to $\operatorname{plim} MC_i(r) = MC_i^*(r)$ for least squares models and, also, given mild regularity conditions for other models as well. Similarly, the product of consistent

random variables must also converge in probability and the inverse of a consistent random variable must also converge in probability. Thus plim $L(r) = L^*(r)$. A final application of Slutsky's theorem then gives the result. $\square$

3.6 Bootstrapped Confidence Intervals

Given the nonlinearity of the proportional value, analytic characterization of the sample distribution of a proportional marginal decomposition will be difficult. Bootstrap approaches to determining confidence intervals are straightforward. Here is one of the simplest. Let $y = \Theta(X) + \epsilon$ with $\epsilon \sim D(0, \Sigma, \ldots)$, where y is $(n \times 1)$ and D is a density function. D could be gaussian, in which case the error covariance matrix is sufficient for identification. A set of bootstrapped vectors $y'_i, i = 1, 2, \ldots, m$ is then created by setting $y'_i = y + \epsilon_i$ with ϵ_i $(n \times 1)$ and distributed $\epsilon_i \sim D(0, \Sigma, \ldots)$. Confidence intervals are then based on the distribution of decomposition estimates based on the models $\{y'_i = \Theta(X)\}_{i=1}^m$. This approach can, of course, be applied to other measures of relative importance.

4 General Properties

This sections develops some general properties of the three measures of relative importance studied in this paper. Of particular importance are the admissibility of these estimators according to the criteria developed in Section 2.

4.1 Proportional Marginal Decomposition

Proportional marginal decomposition takes a very intuitive form in two variable models (with no intercept). Joint explanatory power is allocated to the explanatory variables in proportion to their individual marginal contributions.

$$\varphi_i(w) = \frac{w(\bar{i})}{w(\bar{i}) + w(\bar{j})} \, w(\overline{ij}) \tag{9}$$

Silber, *et al.* (1995) use this simple form of joint allocation of model explanatory power in a study comparing the relative importance of patient and patient characteristics in predicting medical treatment outcomes. In this study individual explanatory factors are aggregated into two groups.

In models with more than two variables the equal proportional change relationship is observed. The proportional change in decomposition share to variable i when variable j is removed from S is equal to the proportional change to j's share when i is removed. Let $\varphi_i(S, w)$ be the decomposition share of variable i when the model is

based on the game w with marginal contributions restricted to the variables in the set S.

$$\frac{\varphi_i(S, w)}{\varphi_i(S \setminus j, w)} = \frac{\varphi_j(S, w)}{\varphi_j(S \setminus i, w)} \tag{10}$$

Feldman (1999, 2002) shows the proportional value has this property. Ortmann (2000), Theorem 2.6 characterizes the proportional value with this property.

4.2 The Averaging Method

The averaging method has the representation

$$AMCV_i(w) = \frac{1}{n!} \sum_{r \in \mathcal{R}(N)} M_{r(i)}(r). \tag{11}$$

Lemma 4.1 $AMCV(w) = Sh(w)$, where $Sh(w)$ is the Shapley value of w.

Proof: The Shapley value of a player in a game is well known to be the player's average marginal contribution over all orders.[8]. $\qquad\square$

Soofi, Retzer and Yasai-Ardekani (2000) show that the averaging method is a maximum entropy estimator. This is also evident from Lemma 6.3, below. The maximum entropy approach is here equivalent to a Bayesian null prior over orderings implying that all orderings should be equally likely. This paper has taken a contrary approach by assuming that the probability of an ordering should be conditioned on the marginal explanatory power of the sets included in the ordering.

4.3 Covariance Decomposition

Covariance decomposition is related to the derivative of model variance with respect to model betas. Model variance $\sigma^2 = \beta'\Sigma\beta$, so $d\sigma^2/d\beta = 2\Sigma\beta$. Thus

$$CVD_i = \beta_i \frac{1}{2} \frac{d\sigma^2}{d\beta_i} = \beta_i \sum_{j=1}^{n} \sigma_{ij}\beta_j. \tag{12}$$

The relationship between covariance decomposition and traditional marginal analysis makes it useful in the marginal analysis, for example, of portfolio positions (See, *e.g.* Litterman (1996).)

Lemma 4.2 *Define* $w(S) = \beta'_S\Sigma\beta$. *Then* $CVD(w) = Sh(w)$.

---

[8]See, for example, Myerson (1991), pp. 438

Proof: Note first that $M_i(S) = \beta_i(\sigma_i^2 + 2\sum_{j \in S, j \neq i} \beta_j \sigma_{ij})$. Variable i is in every S by definition. Each j is clearly in half of the S that contain i (and is 'behind' of i in half of the orderings). Thus, the average marginal contribution is $\beta_i(\sum_{j \in N} \beta_j \sigma_{ij})$. $\qquad\square$

There is another relationship between covariance decomposition and cooperative game theory. Aumann-Shapley prices (Billera and Heath (1986)) are a type of average marginal price based on the value of a nonatomic game. In the context of model performance we can think of the price of variable i as the marginal increase model risk to a small increase in the true value β_i. If model explained variance is the measure of model performance than Aumann-Shapely prices are defined by the relation

$$ p_i = \int_0^1 D^i(r\beta) \ dr, $$

where $D^i = d\sigma^2/d\beta_i$ and the dependence on $r\beta$ is immaterial for a linear derivative. The Aumann-Shapely price vector is then $1/2MC$. The total risk 'price' of each factor is then its magnitude times its price. This is its covariance decomposition component, as shown by equation (12). This relationship may be interpreted to mean that covariance decomposition provides a consistent decomposition of model performance based on linear risk prices.

4.4 Admissibility

The admissibility of relative importance measures is examined criterion by criterion. Only PMD satisfies all admissibility criteria.

4.4.1 Nonnegativity

Lemma 4.3 *The proportional marginal and averaging methods, and any method that can be represented in random order form as in equation (4), satisfy nonnegativity for any monotonic measure of model performance.*

Proof: Any expectation over a set with nonnegative values must be nonnegative. $\qquad\square$

Lemma 4.4 *Covariance decomposition violates the nonnegativity criterion.*

Proof: Consider equation (12) and a two variable example. Then $CVD_i(w) < 0$ if $\beta_2 \sigma_{12} < -\beta_1 \sigma_1^2$. This condition is clearly possible. $\qquad\square$

4.4.2 Proper Exclusion

The proportional marginal method satisfies proper exclusion by design. Covariance decomposition also clearly satisfies proper exclusion, since by equation (12), $\beta_i^* = 0$ requires that $CVD_i = 0$ as well.

Lemma 4.5 *The averaging method violates proper exclusion.*

Proof: Consider equation (11) and a model where $M_1(r) = 0$ for every r such that $r_1 = i$ but that there is a least one S such that $M_i(S) > 0$. This implies that there is at least one r^* such that $j = r^*(i)$ and $M_j(r^*) > 0$ and $AMCV_i(w) > 0$. $\qquad\square$

4.4.3   Proper Inclusion

Lemma 4.6 *Any relative importance measure with a representation as a random order expectation as in equation (4) satisfies proper inclusion.*

Proof: Let r be an ordering such that $r_1 = i$. Since $\mathbf{p}(r|\beta^*) = 0$ (in the limit as defined by equation (8)) if and only if $\beta_i^* = 0$, $E_{\mathbf{p}}M_{r(i)} > 0$ and $E_{\mathbf{p}}[M_{r(i)}(r)] > 0$. $\qquad\square$

Lemma 4.7 *Covariance decomposition does not satisfy proper inclusion.*

Proof: Consider equation (12). Whenever $d\sigma^2/d\beta_i = 0$ then CVD_i must be equal to zero as well. So it can be that $\beta_{i}^* \neq 0$ and $CVD_i = 0$. For example, in a two variable model $d\sigma^2/d\beta_1 = 0$ if $\beta_2\sigma_{12} = -\beta_1\sigma_1^2$ $\qquad\square$

An explanatory variable will receive a zero covariance share whenever it functions as a perfect hedge against the aggregate variance components of other variables that it is correlated with. This result is consistent with marginal analysis, however it is not consistent with what should be expected of a measure of relative importance.

4.4.4   Full Contribution

Proportional marginal, averaging, and covariance decompositions all satisfy the full contribution criterion. The proof that the proportional marginal and averaging methods satisfy full contribution follows from the first proof following, a general proof that applies to any decomposition that can be represented in random order form as in equation (4). Covariance decomposition is addressed in the second proof.

Lemma 4.8 *Let Θ be a model with performance measure μ and explanatory variables N. Construct the statistical cooperative game w based on Θ, N, and μ. Let $R \subset N$ be a set of variables such that for all $i \in R$ and $j \notin R$, $\sigma_{ij} = 0$. Let $\mathbf{p}$ be any probability distribution over $\mathcal{R}(N)$ and let ϕ be the expectation with respect to $\mathbf{p}$ defined by equation (4). Then*

$$\sum_{i \in R} \phi_i(w) = w(R).$$

Proof: The marginal contribution to model performance of any set S must be equal to the sum of the marginal contribution of the sets $S \cap R$ and $S \cap (N \setminus R)$. Thus, the marginal contribution of any variable $i \in S$ to model performance will be $M_i(S^r_{r(i)} \cap S)$. Then for any $i \in S$

$$\phi_i(w) = \sum_{r \in \mathcal{R}(N)} \mathbf{p}(r) M_i(S^r_{r(i)} \cap R)$$

and the sum of $\phi_i(w)$ over all $i \in R$ is necessarily $w(R)$. $\qquad\square$

Lemma 4.9 *Let Θ be a model with performance measure μ and explanatory variables N. Construct the statistical cooperative game w based on Θ, N, and μ. Let $S \subset N$ be a set of variables such that for all $i \in S$ and $j \notin S$, $\sigma_{ij} = 0$. Then*

$$\sum_{i \in S} CVD_i(w) = w(S).$$

Proof: The sum on the left-hand side of the equation in the statement of the lemma expands to $\sum_{i \in S} \sum_{j \in N} \beta_i \sigma_{ij} \beta_j$. But since $\sigma_{ij} = 0$ for $j \notin S$, the formula simplifies to $\sum_{i \in S} \sum_{j \in S} \beta_i \sigma_{ij} \beta_j = \beta'_S \Sigma_S \beta_s = w(S)$. $\qquad\square$

5   Applications and Examples

This section considers the practical application of proportional marginal decomposition. Before examining several examples, the signing of nonnegative relative importance measures by the sign of the associated factor beta is discussed. Only elementary examples are presented in order to focus on basic properties.

All examples are based on OLS regression and explained variance as the measure of model performance. The first three examples are based on simulated data. Example 1 addresses the most basic point that statistical significance is not a reliable indicator of relative importance. Example 2 illustrates the principal limitation of the averaging approach, which is that variables with no statistical significance can have large variance shares. Example 3 illustrates the principal problem of covariance decomposition: statistically significant variables can have zero or negative variance components. In each example PMVD provides the qualitatively correct relationship while one of the other methods fail to do so. Example 4 is based on real data and profiles the performance of a hedge fund against conventional asset benchmarks. It is both an example of the potential practical benefit of PMD and an examination of empirical properties of relative importance measures.

5.1 The Signing Convention

If a relative importance measure satisfies the nonnegativity criterion there is the option to sign importance components by the sign of the factor beta. For simple linear models the signing convention provides useful interpretive information. For some purposes signed PMD components themselves might provide sufficient information. Other nonnegative relative importance measures may be signed as well. In these examples AMCV components are also signed.

5.2 Example 1: Correlated variables

In the first example x_0 is the dependent variable, x_1 and x_2 are correlated independent variables, and x_3 is an uncorrelated independent variable. The correlation between x_1 and x_2 is set to 75%. Table 1 presents the results. In this example, the theoretical joint contribution explained variance of x_1 and x_2 is 77.78% of explained R^2. All measures of relative importance imply that x_1 and x_2 jointly account for about 80% of explained R^2. Casual interpretation of the t-statistics or marginal contributions to explained variance might be interpreted to mean that x_3 is the most important variable.

5.3 Example 2: Spurious independent variable

Table 2 shows the results of an analysis where $x_0 = 1.5x_1 + \epsilon_1$, and $x_2 = x_1 + \epsilon_2$. So x_2 is a spurious regressor. Statistical testing easily determines that x_2 is not statistically significant. PMVD gives it 0.2% variance share. Covariance decomposition also gives x_3 a fairly small share of explained variance, 3.6%. The averaging method, however, allocates x_2 32.0% of the total 79.3% explained variance. This result is an empirical demonstration that the averaging method violates the proper exclusion criterion of relative importance.

In Table 3, $x_0 = 1.5x_1 + 2x_2\epsilon_1$, while $x_3 = x_1 + 0.75x_2 + \epsilon_2$. This time x_3 is the spurious variable. The correlation of x_3 with x_0 is considerably higher than any other variable. Again, the t-test indicates no statistical significance and PMVD gives a zero variance share ($\varphi_3(w) \approx 10^{-6}$). This time the averaging method gives x_3 the largest variance share. In addition to strongly biasing the relative importance of x_1 and x_2 in relation to x_3, the averaging method appears to distort the importance of x_1 relative to x_2 as well.

5.4 Example 3: Negatively correlated variables

Table 4 presents an example based on the relation $x_0 = x_1 + x_2 + \epsilon_1$ with $x_2 = -x_1 + \epsilon_2$. The variances of x_1, ϵ_1 and ϵ_2 are set so that x_1 is uncorrelated with x_0. Note that x_1 is still a statistically significant regressor. All methods except covariance decomposition

indicate that x_1 is an important variable in the model. The theoretical covariance decomposition value is zero because $d\sigma^2/d\beta_1 = 0$. (See equation (12)). This violates the positive inclusion criterion for relative importance measures. The covariance decomposition tells us that x_1 is hedging some of the variance in x_2. This information appears not to be useful for a measure of relative importance.

Table 5 shows that covariance decomposition components can be negative, in violation of the nonnegativity criterion. The basic model is $x_0 = x_1 + x_2 + \epsilon_1$, where $x_2 = -1.5x_1 + \epsilon_2$. The error disturbances are var(x_1)=1, var(ϵ_1)=0.25 and var(ϵ_2)=0.5. Now the CVD variance share for x_1 is -86.0%. The negative value is not from the signing convention of Section 5.1 and has no meaning in the context of a measure of relative importance.

The only reasonable interpretation is as if the model were a portfolio and the factors are portfolio holdings. The portfolio interpretation is that adding x_1 to the portfolio reduces total risk. But x_1 may not be a hedge and, again from the portfolio point of view, most of the diversification benefit of having both x_1 and x_2 in a portfolio is being awarded to x_1. The ratio of PMVD component shares has value $\varphi_1/\varphi_2 = .39$ but the ratio of CVD components has value $CVD_1/CVD_2 = .54$. Covariance decomposition allocates considerably more relative importance to x_1 than PMVD.

5.5 Example 4: Analysis of a Long/Short Fund

This example is a based on a simple factor model analysis of the Laudus Rosenberg Long/Short Fund using conventional asset indexes as explanatory variables. This model might be used to assess the profile of the fund against conventional asset markets. The benchmarks, their abbreviations as will be used here, and their descriptive statistic are found in Table 6. This analysis is based on 81 months of data from January 1998 to September 2004.

5.5.1 Basic analysis

The fund is a small cap long/short fund that reports its holdings. It is approximately market neutral. Table 7 presents the results of a linear OLS regression of the funds returns on the returns of the independent variables shown in Table 6. Table 6 reports the Bera-Jarque nonnormality test results for the residuals. The hypothesis that they are normally distributed cannot be rejected at conventional significance levels.

Review of the results presented in Table 7 reveal a number of patterns. All relative importance measures agree that the TBILL is of negligible importance in spite of its large beta and statistical significance levels. This is logical considering cash volatility and the small likelihood that fund strategy relies on cash derivatives. Small cap growth (SCG) has the largest beta besides cash, the highest statistical significance level and the largest relative importance shares according to all estimators. These results are all consistent with the fund's strategy.

Perhaps the most striking result in this example is that PMVD (and AMCV) find that large cap growth (LCG) is the second most important factor in the model and, particularly, is more important than the mid cap growth factor (MCG). LCG does not appear to be very important if one were to judge from its t-statistic of -1.23 (p=0.221), yet its PMVD share of 9.2% is larger than the 5.8% for MCG. The t-statistic for MCG is 2.50 (p=0.015). More importantly, given that this fund invests in small to (low) mid cap equities, there is no reason to expect strong large cap factor loadings. Thus, it would be logical to suspect that the LCG PMVD component is not accurate.

Large cap value (LCV) also has a sizable negative PMVD component. This sign agreement between LCG and LCV is in contrast to the sign reversal found for mid and small cap indexes. This later pattern suggests that a significant aspect the fund's strategy might be to be long the small cap value premium and short the mid cap value premium. The sign agreement for large cap factors suggest that the correlation between these factors might provide a partial explanation for their observed levels of statistical significance.

Pursuing the hypothesis that there might be a common large cap factor, a simple test would be to constrain the betas of the large cap growth and value factors to be equal, or almost equivalently, to replace them with a large cap factor represented by the S&P 500. If this is done the R^2 decreases from 55.52% to 55.19%. The F-test on the constraint has value 0.523, with $p = 0.47$, indicating the hypothesis that the coefficients are the same cannot be rejected. The t-statistic for the large cap factor increases to -2.11 (p= 0.038) and the PMVD component is -16.6%. This represents a slight increase from the combined -15.6% of the two factors separately. (Also, the MCG PMVD component rises slightly from 5.8% to 6.4% and its t-statistic rises to 2.59. (p=0.011).)

Also noteworthy is that the AMCV large cap component is only -10.6%, in comparison to the -14.5% sum of the individual components. These results are consistent with the known properties of the Shapely value. If players in a game are aggregated, there aggregate value may be very different from the sum of their individual values. The proportional value is much less subject to this type of aggregation bias.[9] Further results of this constraint (or aggregation) test are not reported to save space.

The constraint test shows that the aggregate large cap factor is statistically significant. This result and the stability of the total large cap PMVD share provides credibility for reliance on PMD measures of relative importance. It further illustrates the potential value of a measure of relative importance in providing direction for model analysis.

---

[9]Note that there are two types of aggregation that might be of interest. The first is linearity or additivity, *e.g.* $\phi(v+w) = \phi(v) + \phi(w)$. The Shapley value is additive while the proportional value is not. The aggregation considered here, of variables (players) must be considered of more importance in statistical modeling.

5.5.2 Bootstrap analysis

Bootstrap simulation provides another way to examine the results and the reliability of relative importance measures. A simple bootstrap with 500 simulations generated as described in Section 3.6 was constructed using normally distributed bootstrap residuals. Using normally distributed bootstrap residuals is a reasonable choice given the approximate normality of the actual residuals. It is also a desirable choice so that the characteristics of the bootstrapped distributions of relative importance measures can be clearly associated with the measures and not the residual distribution.

Summary statistics from bootstrap simulations are presented in Table 8. Mean bootstrap betas and their standard deviations are consistent with the OLS results. Bootstrap PMVD components tend to be slightly larger in magnitude than the empirical shares displayed in Table 7. This is a consequence of the reported skew in component shares. AMCV component means tend to be smaller than the empirical results. CVD bootstrap means appear to be quite close to empirical values. All relative importance measures have large normality rejection rates based on the Bera-Jarque test. The effect of the high levels of skew and kurtosis in relative importance measures is difficult to evaluate simply from summary statistics.

Figure 1 shows the univariate distributions of PMVD component shares for all factors except TBILL, which is excluded to save space. It is apparent that there are three basic types of distributions: approximately normal distributions such as observed for SCG; highly skewed distributions that look almost exponential in nature such as observed for LCG, and symmetric highly kurtotic distributions such as observed for HYLD. The skew and kurtosis appear to result from the highly nonlinear implied ordering probabilities necessary for proportional decompositions to satisfy limit proper exclusion. The effect is perhaps most dramatic for factors that have strong influence but little marginal contribution, such as LCG, where the skew dominates. Factors with little explanatory power at any point in the relative importance ordering, like HYLD, appear to have simply the high kurtosis. The skew and kurtosis indicate that these components must be assumed to have less reliability than factors like SCG. Note, also, that all factors with reasonably large PMVD shares (say, greater than about 3%) have small tails extending across into the quadrant of the opposite sign. Factors with large PMVD shares are unlikely to have bootstrap realizations with significant component shares of opposite magnitude.

Figure 2 shows univariate AMCV component share distributions. The striking difference with the PMVD distributions is their bimodality. The bimodality results from the combination of a factor's low statistical significance with the averaging method's failure to satisfy proper exclusion.

Figure 3 is provided for completeness and shows univariate CVD component shares. A tendency for extended tails is evident in these distributions as well. This is probably a reflection of the fact that this measure is a sum of weighted pairwise beta products. These products will be distributed approximately according to the χ^2 distribution.

Figure 4 shows bivariate scatter plots of the distribution of bootstrap betas. This figure is provided for reference. The generally elliptical nature of these scatters and their slopes are consistent with what we should expect.

Figure 5 shows bivariate scatter plots of the distribution of bootstrap PMVD component shares. These results are very striking. Only a limited analysis can be presented here. I will focus on the LCG and LCV scatters with other factors. Notice their generally right-triangular form. The LCG x LT chart (LCG on the x-axis) is easiest to interpret. Scaled LCG values are always larger than scaled LT values (see chart caption for scaling definition). This profile is also found with SCV, both positive factors. The more typical profile for scatters with LCG and LCV is the upper triangular profile that can be seen for LCG x LCV and with other negative factors. The upper triangular profile is described in the caption for the figure. The interpretation in both cases is the same. Low scaled values of LCG correspond to large negative shares. As LCG or MCV share increases, the shares of all other factors are constrained (not so surely for MCG, however). This might be thought to be due to the overall constraint that shares must add to total R^2. But if this were true, the same pattern would be observed for all factor combinations, and, particularly, for SCG with it large component shares.

Figure 6 shows bivariate scatter plots of the distribution of AMCV component shares. The bimodality found in the univariate distributions of Table 2 is now reflected as a clover structure when at least one factor is bimodal. Finally, for completeness, Figure 7 shows CVD bootstrap component bivariate scatters. These scatters appear very similar to the beta scatters of Figure 4.

6   Discussion

6.1   Generalized Proportional Marginal Decomposition

The axiom of equal proportional effect could be relaxed to require only proportional effect:

Axiom 6.1 PROPORTIONAL EFFECT: $\left| \dfrac{\partial \ln L(r)}{\partial \ln w(S)} \right| = \alpha, \quad \alpha \in \mathbb{R}$

Note that axiom of limit proper exclusion is formally satisfied whenever $\alpha > 0$. It is then straightforward to verify the following lemma.

Lemma 6.1 *Given a model Θ, a performance measure μ and a game w generated by μ, the likelihood determined by the axioms of anonymity, limit proper exclusion and equal proportional effect is*

$$L(r, \alpha) = \left(\prod_{S \in r} w(S) \right)^{-\alpha}, \quad \alpha > 0.$$

Defining a generalized normalizing factor $P(N, \alpha)$

$$P(N, \alpha) = \left(\sum_{r \in \mathcal{R}(N)} L(r, \alpha) \right)^{-1}, \tag{13}$$

a continuum of probability distributions indexed by α results:

$$\mathbf{p}^\alpha(r) = P(N, \alpha)L(r, \alpha). \tag{14}$$

Substitution in the expectation (4) then leads to the following result:

Lemma 6.2 *The expected contribution of variable i to the performance of model Θ relative to the game w defined by a performance measure μ and according to the probability distribution defined by the axioms of anonymity, limit proper exclusion and proportional effect is*

$$\varphi_i^\alpha(w) = P(N, \alpha) \sum_{r \in \mathcal{R}(N)} L(r, \alpha) M_{r(i)}(r),$$

for $\alpha > 0$.

Since $L(r, \alpha = 0) = 1$, the Shapley value is the limit point for this continuum:

Lemma 6.3 $\varphi^{\alpha=0}(w) = Sh(w)$.

Finally, the next lemma is follows from pervious results in Section 4.4.

Lemma 6.4 *Each element in the continuum of generalized proportional marginal decompositions identified by Lemma 6.2 is an admissible measure of relative importance.*

Proof: Every element of this continuum clearly satisfies nonnegativity and proper exclusion. Lemma 4.6 guarantees proper inclusion and Lemma 4.9 proves that $\varphi^\alpha(w)$ satisfies full contribution. $\qquad\square$

Note that $P(N, \alpha)$ is only a potential when $\alpha = 1$. This is in the sense that $\varphi_i^\alpha(w) \neq P(N, \alpha)/P(N \setminus i, \alpha)$ for $\alpha \neq 1$. Lemma 3.2 does not generalize.

6.2 Calibration

The existence of the generalized continuum of proportional marginal decompositions as described by Lemma 6.2 invites calibration. Why not find the empirical α that best fits any particular model? After all, relative importance is defined by reference to a probability distribution over orderings of model explanatory variables.

For example, consider a bivariate OLS model. What is the relationship between the probability that one variable is "most important" and the probability that it's marginal contribution to variance is greatest? According to the basic proportional marginal decomposition approach, the probability that one variable is most important is proportional to its variance share. Lemma 6.2 implies that this probability could be matched precisely by relaxing the requirement of equal proportionality.

Calibration may be useful problems indirectly related to relative importance, but I believe that calibration is not a useful approach to the assessment of relative importance. Such an approach would make relative importance a function of the number of observations and quantity of unexplained variance (*i.e.* noise) in the model. The seems unhelpful. Further, these considerations suggest an additional potential criterion for relative importance. Theoretical relative importance shares should be invariant with respect to noise and the number of observations. Note this criterion would reject only calibration and not estimation of relative importance using a generalized proportional decomposition with fixed $\alpha \neq 1$.

6.3 Relationship to Cooperative Game Theory

The properties of the cooperative value functions associated with different measures of relative importance provide insight into their mechanism. The Shapley value is linear while the proportional value is inherently nonlinear. The proportional value has the property that a player with zero individual worth must receive zero value.[10] If there are only two explanatory variables, the Shapley value implies equal division of joint explanatory power while the proportional value requires that it is divided in proportion to individual marginal explanatory power. Noncooperative models of the Shapley value give players equal bargaining power while noncooperative models of the proportional value give players with greater expected payoffs more bargaining power[11] The inadmissibility of two types of linear estimators of relative importance and the admissibility of a continuum of proportional estimators suggest that relative importance is an inherently nonlinear measure.

7 Conclusion

This paper advances four fundamental criteria for the admissibility of measures of statistical relative importance. The measure of relative importance developed here, proportional marginal decomposition, is admissible under these criteria. PMD and the generalization developed in Section 6.1 are the only relative importance measures know to be admissible.

---

[10]Assuming all other players have positive individual worth, see Feldman (2002), Theorem 6.1.

[11]This statement assumes that bargaining power is equivalent to the probability of having the opportunity to make a proposal. Regarding models of the Shapley value, see Gul (1986) or Hart and Mas-Colell (1989). Feldman (2002) proposes a noncooperative implementation of the proportional value.

Proportional marginal decompositions may be estimated consistently. Examples demonstrate the practical value of this approach over several existing alternative measures. The empirical example also demonstrate that the nonlinearity of the PMD method calls for some caution.

The results developed in this paper and much of the relative importance literature cited demonstrate the close relationship between measures of relative importance and cooperative value. Feldman (2002) advances a dual model of cooperative value in which the Shapley and proportional values are merely *modes* of manifestation a more fundamental notion of value. The dual model proposes that one mode might be found more relevant in any specific situation. This paper demonstrates that the proportional mode provides a better representation of statistical relative importance.

Appendix A: Computing the Proportional Value

Calculating a proportional marginal decomposition using the expectations approach quickly becomes impractical from a computational viewpoint. A model with 10 independent variables would require the evaluation of 10! orderings. Feldman (1999) and Ortmann (2000) show the proportional value may be computed using a ratio potential, which is defined by the recursive relation

$$P(S) = w(S) \left(\sum_{i \in S} P(S \setminus i)^{-1} \right)^{-1}, \tag{15}$$

for all $S \subset N$, and $P(\emptyset) = c > 0$. The proportional value of player i is then $\varphi_i(w) = P(N)/P(N \setminus i)$.

Appendix B: Proof of Lemma 3.2

Proof: The structure of $P(N \setminus i)$ is identical to $P(N)$ except that it is based on the permutation set $\mathcal{R}(N \setminus i)$. Consider any $r^o \in \mathcal{R}(N \setminus i)$. There are n orderings $(r^1, r^2, \ldots, r^n) \in \mathcal{R}(N)$ that result when i is inserted into r^0 at the jth position so that $r_j^j = i$. If

$$L(r^o) = \sum_{j=1}^{n} L(r^j) M_{r(i)}(r^j), \tag{16}$$

the result follows. The key to this reduction becomes more apparent when the marginal contribution $M_{r(i)}(r)$ is represented as $w(S_j^j) - w(S_{j-1}^j)$, that is, as the difference between the worth of smallest set including i in the order j and the worth of the just smaller set.

The relation on the right hand of the above equation can then be expanded into $2n - 1$ separate terms. Consider first the term with $w(S_n^n)$ in the numerator. Since $S_n^n = N$, this term simplifies to the term on the left side of the equation. The remaining terms cancel each other out. Specifically, consider the terms associated with S_{j-1}^{j-1} and S_{j-1}^j:

$$\frac{w(S_{j-1}^{j-1})}{w(S_1^{j-1}) \cdots w(S_{j-1}^{j-1}) \cdots w(S_n^{j-1})} + \frac{-w(S_{j-1}^j)}{w(S_1^j) \cdots w(S_{j-1}^j) \cdots w(S_n^j)} = 0. \qquad (17)$$

The sum is zero because for all $k < j - 1$ and all $k > j - 1$, $w(S_k^{j-1}) = w(S_k^j)$ because either i has not yet entered, or i has already entered. For position $j - 1$, where $S_{j-1}^{j-1} \neq S_{j-1}^j$ these terms cancel out on both sides. $\qquad \square$

References

BILLERA, L. J. AND D. C. HEATH (1982): "Allocation of shared cost: A set of axioms yielding a unique procedure," *Mathematics of Operations Research*, **7**:32-39.

CAMPBELL, JOHN H. (1992): "Explaining the variance of price-dividend ratios," *Review of Financial Studies*, **5**:243-280.

CHEVAN, ALBERT AND MICHAEL SUTHERLAND (1991): "Hierarchical partitioning," *The American Statistician*, **45**:90-96.

EHRENBERG, A. S. C. (1990): "The unimportance of relative importance," *American Statistician*, **44**:260.

FELDMAN, BARRY (1999): *The Proportional Value of a Cooperative Game*, `netec.mcc.ac.uk/WoPEc/data/Papers/ecmwc20001140.html`.

——— (2002) *A Dual Theory of Cooperative Value*, `papers.ssrn.com/abstract=317284`.

FIRTH, DAVID (1998): *Relative Importance of Explanatory Variables*, prepared for *Statistical Issues in the Social Sciences*, Stockholm, October 1998, Oxford: Nuffield College, `www.nuff.ox.ac.uk/sociology/alcd/relimp.pdf`.

FREES, EDWARD W. (1998): "Relative importance of risk sources in insurance systems," *North American Actuarial Journal*, **2**:34-48. Comments pp. 48-54.

GOLDBERGER, ARTHUR M. AND CHARLES F. MANSKI (1995): "*The Bell Curve* by Hernstein and Murray, *Journal of Economic Literature*, **33**:762-776.

GUL, FARUK (1989): "Bargaining foundations of the Shapley value," *Econometrica*, **57**:81-95.

HART, SERGIU, AND ANDREU MAS-COLELL (1989): "Potential, value and consistency," *Econometrica*, **57**:589-614.

——— AND ——— (1996): "Bargaining and value," *Econometrica*, **64**:357-380.

HEALY, M. J. R. (1990): "Measuring importance," *Statistics in Medicine*, **9**:633-637.

HECKMAN, JAMES M. (1995): "Lessons from the bell curve," *Journal of Political Economy*, **103**:1091-1120.

HOOKER, R. H. AND G. U. YULE (1906): "Note on estimating the relative influence of two variables upon a third," *Journal of the Royal Statistical Society*, **69**:197-200.

KING, GARY (1987): "How not to lie with statistics: Avoiding common mistakes in quantitative political science," *American Journal of Political Science*, **30**:666-687.

KRUSKAL, WILLIAM H. (1987): "Relative importance by averaging over orderings," *American Statistician*, **41**:6-10.

—— AND RUTH MAJORS (1989): "Concepts of relative importance in scientific literature," *American Statistician*, **43**:2-6.

LINDEMAN, RICHARD. H., PETER F. MERENDA, AND RUTH Z. GOLD (1980): *Introduction to Bivariate and Multivariate Analysis*, Scott Foresman, Glenview, IL.

MYERSON, ROGER (1991): *Game Theory: Analysis of Conflict*, Cambridge: Harvard University Press.

ORTMANN, K. M. (2000): "The proportional value for positive cooperative games," *Mathematical Methods of Operations Research*, **51**:235-48.

PRATT, JOHN W. (1987): "Dividing the indivisible: Using simple symmetry to partition variance explained," In Pukkila, T. and Puntanen, S., editors, *Proceedings of the Second International Tampere Conference in Statistics*, pages 245-260.

SILBER, JEFFREY H., PAUL R. ROSENBAUM, AND RICHARD N. ROSS (1995): "Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics?," *Journal of the American Statistical Association*, **90**:7-18.

SCHEMPER, M. (1993): "The relative importance of prognostic factors in studies of survival," *Statistics in Medicine*, **12**:2377-2382.

SHAPLEY, LLOYD S. (1953): "Additive and Non-Additive Set Functions," Ph.D. Thesis, Princeton: Princeton University.

SIMS, CHRISTOPHER A. (1980): "Macroeconomics and reality," *Econometrica*, **48**:1-48.

SOOFI, EHSAN S., JOSEPH J. RETZER, AND MASOUD YASAI-ARDEKANI (2000): "A framework for measuring importance of variables with application to management research and decision models," *Decision Sciences*, **31**:1-31.

STUFKEN, JOHN (1992): "On hierarchical partitioning," *American Statistician*, **46**:70-71.

THEIL, HENRI (1971): *Principles of Econometrics*, Wiley.

—— (1987): "How many bits of information does an independent variable yield in a multiple regression?," *Statistics and Probability Letters*, **6**:107-108.

—— AND C. CHUNG (1988): "Information-theoretic measures of fit for univariate and multivariate linear regressions," *American Statistician*, **42**:249-252.

Table 1: Two 75% correlated and one uncorrelated independent variables

Parameter	beta	std. err.	t-stat	p-val	MV	PMVD	AMCV	CVD
Intercept	-0.0439	0.1309	-0.34	0.7377				
x_1	1.1138	0.1971	5.65	0.0000	0.0531	21.2%	22.4%	22.2%
x_2	1.2031	0.2042	5.89	0.0000	0.0577	24.9%	23.6%	23.7%
x_3	1.1849	0.1349	8.78	0.0000	0.1282	13.0%	13.1%	13.2%

The true model is $x_0 = x_1 + x_2 + x_3 + \epsilon_1$, var($x_1$)= var($x_2$)=var($x_3$)=1.0, $\sigma_{12} = 0.75$. var(ϵ_1) = 2.0. n = 250. R^2= 0.5911.

MV is the marginal contribution to variance, PMVD is proportional marginal variance decomposition, AMCV is average marginal contribution to variance, and CVD is the covariance decomposition. Marginal variance contributions are not calculated for the intercept. Variance decompositions do not consider the contribution of the intercept term.

Correlations	x_0	x_1	x_2	x_3
x_0	100.0%	63.0%	64.5%	36.7%
x_1	63.0%	100.0%	75.7%	-0.4%
x_2	64.5%	75.7%	100.0%	2.8%
x_3	36.7%	-0.4%	2.8%	100.0%

In this example, the correlation between x_1 and x_2 is 75%. T-statistics correctly indicate that x_3 is much more precisely estimated. All variance decomposition methods correctly imply that x_1 and x_2 are together by far more important in driving model performance.

Table 2: Spurious variable in model with one true independent variable

Parameter	beta	std. err.	t-stat	p-val	MV	PMVD	AMCV	CVD
Intercept	0.0169	0.0277	0.61	0.542				
x_1	1.4720	0.1220	12.06	0.000	0.1537	79.1%	47.3%	75.7%
x_2	0.0677	0.1051	0.64	0.520	0.0004	0.2%	32.0%	3.6%

The true model is $x_0 = 1.5x_1 + \epsilon_1$, with $x_2 = x_1 + \epsilon_2$ and $\text{var}(x_1) = 0.25, \text{var}(\epsilon_1) = 0.16$, and $\text{var}(\epsilon_2) = 0.0625$. n = 200. $R^2 = 0.793$.

MV is the marginal contribution to variance, PMVD is proportional marginal variance decomposition, AMCV is average marginal contribution to variance, and CVD is the covariance decomposition. Marginal variance contributions are not calculated for the intercept. Variance decompositions do not consider the contribution of the intercept term.

Correlations	x_0	x_1	x_2
x_0	100.0%	88.5%	79.1%
x_1	88.5%	100.0%	91.5%
x_2	79.1%	91.5%	100.0%

In this simple example the variable x_2 is not part of the true model in spite of its relatively high correlation with x_0. All measures except AMCV, the averaging method, correctly identify this condition.

Table 3: Spurious variable in model with two true independent variables

Parameter	beta	std. err.	t-stat	p-val	MV	PMVD	AMCV	CVD
Intercept	-0.0328	0.0333	-0.98	0.326				
x_1	1.5635	0.2248	6.95	0.000	0.0520	30.1%	13.4%	22.1%
x_2	2.1296	0.1779	11.97	0.000	0.1540	48.9%	31.9%	56.9%
x_3	-0.0017	0.2159	-0.01	0.994	0.0000	0.0%	33.6%	-0.1%

The true model is $x_0 = 1.5x_1 + 2x_2 + \epsilon_1$, with $x_3 = x_1 + 0.75x_2 + \epsilon_2$ and $\mathrm{var}(x_1) = 0.25$, $\mathrm{var}(\epsilon_1) = 0.25$, and $\mathrm{var}(\epsilon_2) = 0.16$. n = 200. $R^2 = 0.789$.

MV is the marginal contribution to variance, PMVD is proportional marginal variance decomposition, AMCV is average marginal contribution to variance, and CVD is the covariance decomposition. Marginal variance contributions are not calculated for the intercept. Variance decompositions do not consider the contribution of the intercept term. PMVD and AMCV variance components are signed by the sign of the factor beta.

Correlations	x_0	x_1	x_2	x_3
x_0	100.0%	29.6%	59.8%	73.8%
x_1	29.6%	100.0%	-47.3%	69.4%
x_2	59.8%	-47.3%	100.0%	23.2%
x_3	73.8%	69.4%	23.2%	100.0%

In this case, PMVD allocates almost a zero variance share ($\varphi_3 \approx 10^{-6}$) to x_3 even though x_3 has the highest simple Pearson correlation with x_0. The averaging approach gives x_3 the largest share.

Table 4: Negatively correlated independent variables I

Parameter	beta	std. err.	t-stat	p-val	MV	PMVD	AMCV	CVD
Intercept	0.0182	0.0194	0.94	0.3472				
x_1	0.9880	0.0456	21.66	0.0000	0.5987	33.3%	29.9%	-0.2%
x_2	1.0017	0.0414	24.22	0.0000	0.7486	41.6%	44.9%	75.0%

The true model is $x_0 = x_1 + x_2 + \epsilon_1$, $x_2 = -x_1 + \epsilon_2$. var(x_1)= 1, var(ϵ_1)=0.25, var(ϵ_2)=.5, n = 200. R^2= 0.7486.

MV is the marginal contribution to variance, PMVD is proportional marginal variance decomposition, AMCV is average marginal contribution to variance, and CVD is the covariance decomposition. Marginal variance contributions are not calculated for the intercept. Variance decompositions do not consider the contribution of the intercept term.

Correlations	x_0	x_1	x_2
x_0	100.0%	-0.1%	38.7%
x_1	-0.1%	100.0%	-89.5%
x_2	38.7%	-89.5%	100.0%

In this example x_1 and x_2 are highly negatively correlated. The component of x_2 correlated with x_1 is exactly balanced by x_1. The result is that x_1 has an almost zero covariance decomposition component. Note all methods except covariance decomposition yield similar implications for the relative importance of the independent variables.

Table 5: Negatively correlated independent variables II

Parameter	beta	std. err.	t-stat	p-val	MV	PMVD	AMCV	CVD
Intercept	0.0182	0.0194	0.94	0.3472				
x_1	0.9888	0.0648	15.27	0.0000	0.1679	24.4%	30.2%	-86.0%
x_2	1.0017	0.0414	24.22	0.0000	0.4227	61.4%	55.6%	171.8%

The true model is $x_0 = x_1 + x_2 + \epsilon_1$, $x_2 = -1.5x_1 + \epsilon_2$. var($x_1$)= 1, var($\epsilon_1$)=0.25, var($\epsilon_2$)=.5, n = 200. R^2= 0.8580.

MV is the marginal contribution to variance, PMVD is proportional marginal variance decomposition, AMCV is average marginal contribution to variance, and CVD is the covariance decomposition. Marginal variance contributions are not calculated for the intercept. Variance decompositions do not consider the contribution of the intercept term.

Correlations	x_0	x_1	x_2
x_0	100.0%	-66.0%	83.1%
x_1	-66.0%	100.0%	-94.9%
x_2	83.1%	-94.9%	100.0%

As in the example of Table 4a, x_1 and x_2 are highly negatively correlated. In this example the component of x_2 correlated with x_1 is not completely neutralized by x_1. The result is a negative covariance component for x_1. All methods except covariance decomposition yield similar implications for the relative importance of the independent variables. The negative covariance component reflect the fact that $var(\beta_1 x_1) < var(\beta_1 x_1 + \beta_2 x_2)$, so x_1 could be interpreted as a hedging factor in some contexts.

Table 6: Descriptive Statistics for Long-Short Fund Analysis

Factor	Abbrev.	Mean	Std. Dev.	Skew	Excess Kurto-sis	Bera Jarque p-value
S&P/BARRA 500 Growth	LCG	0.337%	0.055	-0.359	-0.453	0.2960
S&P/BARRA 500 Value	LCV	0.479%	0.050	-0.483	0.683	0.0941
S&P/BARRA MidCap 400 Growth	MCG	1.017%	0.069	-0.126	0.954	0.1935
S&P/BARRA MidCap 400 Value	MCV	0.973%	0.053	-0.323	0.994	0.0933
Russell 2000 Growth	SCG	0.417%	0.084	-0.201	0.125	0.7416
Russell 2000 Value	SCV	0.854%	0.049	-0.791	1.719	0.0001
MSCI EAFE Free	EAFE	0.409%	0.047	-0.433	0.182	0.2663
MSCI Emerging Mkts	EM	0.637%	0.074	-0.856	1.957	0.0000
LB Hi-Yld	HYLD	0.444%	0.025	-0.451	2.149	0.0001
LB U.S. LT Govt	LT	0.658%	0.026	-0.850	2.064	0.0000
U.S. 30 Day TBill	TBILL	0.275%	0.002	0.011	-1.578	0.0150
Laudus Rosenberg Fund		0.269%	0.039	0.249	0.984	0.1286
Residuals		0.000%	0.026	-0.428	0.386	0.2266

The excess kurtosis of the normal distribution is zero. The Bera Jarque p-values are based on the Bera Jarque test statistic and represent the confidence level in rejecting the hypothesis of asset return distribution normality based the sample values for the skew and kurtosis of the distribution. This test statistic is distributed $\chi_{(2)}$. The hypothesis that the residuals are normality distributed cannot be rejected at conventional significance levels.

Corr.	LR	LCG	LCV	MCG	MCV	SCG	SCV	EAFE	EM	HYLD	LT	TBILL
LR	1.00	-0.63	-0.54	-0.56	-0.42	-0.64	-0.43	-0.53	-0.54	-0.36	0.28	-0.05
LCG	-0.63	1.00	0.76	0.81	0.60	0.74	0.52	0.79	0.67	0.41	-0.24	-0.03
LCV	-0.54	0.76	1.00	0.71	0.90	0.62	0.74	0.78	0.72	0.46	-0.30	0.01
MCG	-0.56	0.81	0.71	1.00	0.70	0.90	0.73	0.74	0.72	0.44	-0.21	0.01
MCV	-0.42	0.60	0.90	0.70	1.00	0.59	0.85	0.70	0.70	0.41	-0.19	0.00
SCG	-0.64	0.74	0.62	0.90	0.59	1.00	0.77	0.72	0.75	0.52	-0.22	-0.07
SCV	-0.43	0.52	0.74	0.73	0.85	0.77	1.00	0.66	0.69	0.52	-0.18	-0.09
EAFE	-0.53	0.79	0.78	0.74	0.70	0.72	0.66	1.00	0.74	0.47	-0.23	-0.11
EM	-0.54	0.67	0.72	0.72	0.70	0.75	0.69	0.74	1.00	0.56	-0.29	-0.18
HYLD	-0.36	0.41	0.46	0.44	0.41	0.52	0.52	0.47	0.56	1.00	-0.02	-0.19
LT	0.28	-0.24	-0.30	-0.21	-0.19	-0.22	-0.18	-0.23	-0.29	-0.02	1.00	0.05
TBILL	-0.05	-0.03	0.01	0.01	0.00	-0.07	-0.09	-0.11	-0.18	-0.19	0.05	1.00

Parameter	beta	std. err.	t-stat	p-val	MV	PMVD	AMCV	CVD
Intercept	0.0074	0.0068	1.08	0.2831				
LCG	-0.1829	0.1481	-1.23	0.2210	0.0098	-9.2%	-9.0%	16.0%
LCV	-0.1963	0.2088	-0.94	0.3505	0.0057	-6.4%	-5.5%	13.6%
MCG	0.3563	0.1423	2.50	0.0147	0.0403	5.8%	6.6%	-35.4%
MCV	-0.1965	0.2235	-0.88	0.3823	0.0050	-0.9%	-2.9%	11.2%
SCG	-0.5158	0.1371	-3.76	0.0003	0.0911	-26.7%	-13.8%	71.9%
SCV	0.3578	0.2010	1.78	0.0794	0.0204	3.6%	3.6%	-19.6%
EAFE	0.0895	0.1330	0.67	0.5032	0.0029	0.7%	4.2%	-5.7%
EM	0.0172	0.0843	0.20	0.8386	0.0003	0.1%	4.8%	-1.8%
HYLD	-0.0793	0.1702	-0.47	0.6425	0.0014	-0.1%	-2.0%	1.8%
LT	0.1564	0.1363	1.15	0.2552	0.0085	1.1%	2.2%	2.9%
TBILL	-2.5169	2.1349	-1.18	0.2424	0.0089	-1.0%	-1.0%	0.6%

Acronyms are defined in Table 6. The equation R^2 is 55.52%.

MV is the marginal contribution to variance, PMVD is proportional marginal variance decomposition, AMCV is average marginal contribution to variance, and CVD is the covariance decomposition. Marginal variance contributions are not calculated for the intercept. Variance decompositions do not consider the contribution of the intercept term.

Note low statistical significance level for LCG in comparison to its PMD variance share. Note also, low PMVD allocations compared to AMCV and CVD allocations for EAFE, EM, HYLD, and LT. This result is consistent with statistical significance levels.

Table 8: Bootstrap simulation statistics

Measure	Factor	Mean Value	Std. Dev.	Skew	Excess Kurtosis	Bera-Jarque statistic	p-val
Beta	LCG	-0.176	0.136	0.083	-0.391	3.77	0.1518
	LCV	-0.195	0.198	0.091	-0.247	1.96	0.3758
	MCG	0.355	0.143	-0.086	0.253	1.95	0.3778
	MCV	-0.193	0.218	-0.146	0.092	1.96	0.3762
	SCG	-0.514	0.133	-0.058	-0.017	0.28	0.8688
	SCV	0.362	0.195	0.162	0.023	2.20	0.3335
	EAFE	0.068	0.132	-0.284	0.345	9.18	0.0102
	EM	0.022	0.079	-0.052	-0.172	0.83	0.6588
	HYLD	-0.081	0.171	-0.118	-0.081	1.30	0.5209
	LT	0.154	0.135	-0.070	-0.028	0.43	0.8077
	TBILL	-2.449	2.148	-0.180	-0.121	3.01	0.2220
PMVD	LCG	-0.100	0.098	-1.058	0.800	106.54	0.0000
	LCV	-0.054	0.059	-1.438	2.821	338.21	0.0000
	MCG	0.057	0.021	-0.627	1.537	82.02	0.0000
	MCV	-0.014	0.023	-1.135	4.811	589.52	0.0000
	SCG	-0.242	0.082	-0.169	-0.272	3.93	0.1402
	SCV	0.035	0.022	0.269	-0.598	13.50	0.0012
	EAFE	0.000	0.032	-3.494	15.499	6022.09	0.0000
	EM	-0.003	0.028	-3.642	20.478	9842.13	0.0000
	HYLD	-0.008	0.019	-2.230	6.120	1194.58	0.0000
	LT	0.020	0.025	1.742	4.248	628.61	0.0000
	TBILL	-0.014	0.017	-1.739	4.415	658.04	0.0000
AMCV	LCG	-0.079	0.052	1.691	2.794	401.02	0.0000
	LCV	-0.043	0.043	1.003	0.201	84.74	0.0000
	MCG	0.066	0.021	-5.354	36.949	30830.22	0.0000
	MCV	-0.020	0.027	1.255	0.354	133.77	0.0000
	SCG	-0.136	0.031	-0.302	-0.109	7.85	0.0198
	SCV	0.037	0.016	-3.653	18.360	8134.83	0.0000
	EAFE	0.015	0.047	-1.146	-0.347	112.00	0.0000
	EM	0.001	0.052	-0.523	-1.476	68.14	0.0000
	HYLD	-0.016	0.024	-0.263	-0.428	9.61	0.0082
	LT	0.025	0.022	1.034	2.103	181.33	0.0000
	TBILL	-0.014	0.018	-1.911	5.604	958.80	0.0000
CVD	LCG	0.159	0.128	0.228	-0.327	6.56	0.0375
	LCV	0.142	0.147	0.385	0.520	17.95	0.0001
	MCG	-0.343	0.134	0.119	0.512	6.63	0.0364
	MCV	0.115	0.132	0.612	0.691	41.20	0.0000
	SCG	0.717	0.219	0.323	0.028	8.72	0.0128
	SCV	-0.190	0.100	-0.051	0.119	0.51	0.7742
	EAFE	-0.037	0.083	0.675	0.660	47.08	0.0000
	EM	-0.018	0.079	0.378	0.143	12.34	0.0021
	HYLD	0.025	0.044	0.942	1.078	98.18	0.0000
	LT	0.034	0.033	0.905	0.593	75.53	0.0000
	TBILL	0.012	0.016	2.285	7.090	1482.53	0.0000
	Intercept	0.005	0.007	0.241	-0.051	4.90	0.0865
	R^2	0.587	0.060	-0.234	0.174	5.19	0.0746

Statistics for bootstrap test based on 500 simulations. Mean betas and standard deviations are consistent with OLS results reported in Table 7. PMVD and AMCV share components show high levels skew and kurtosis. CVD share components show significant nonnormality as well.
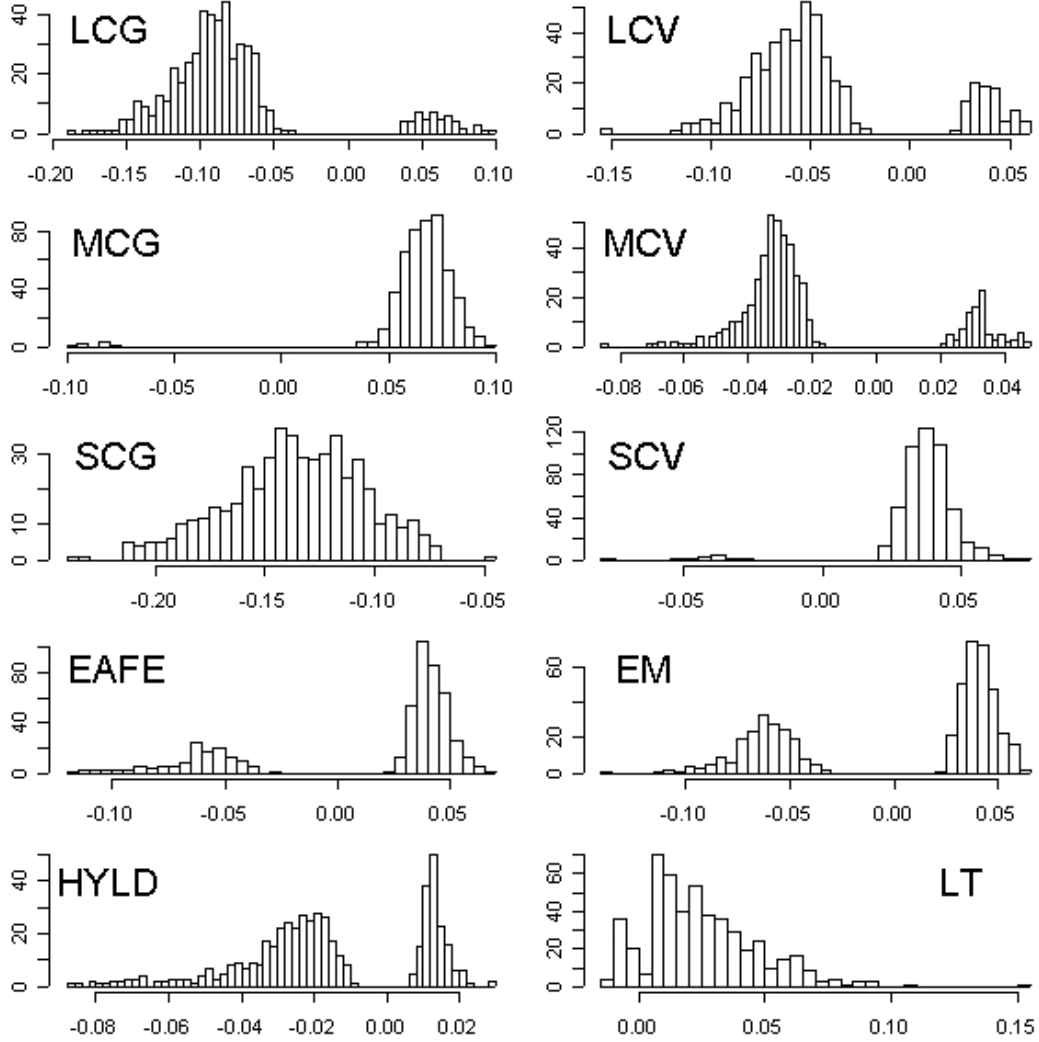
Figure 1: PMVD component bootstrap distributions
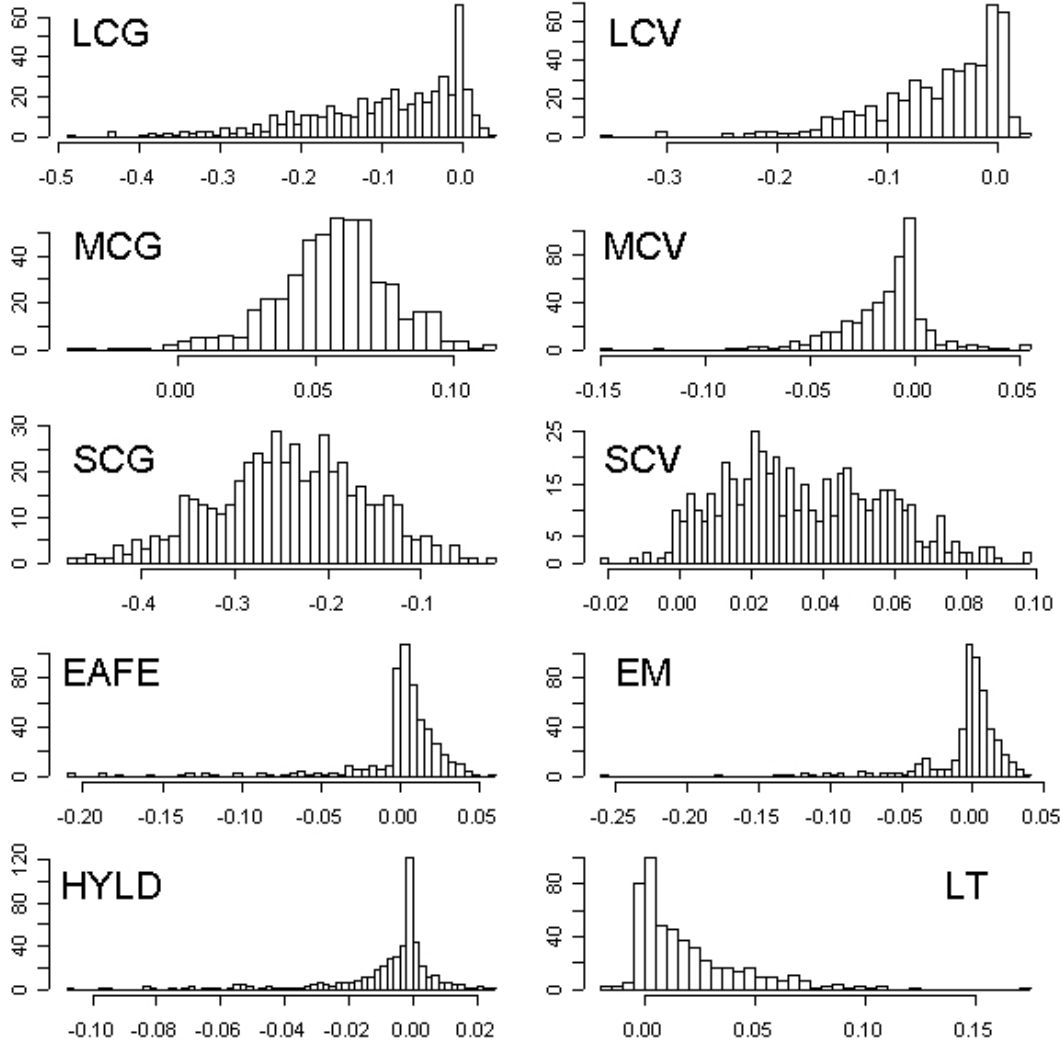
Acronyms are defined in Table 6. PMVD variance components are signed by the sign of the factor beta, see Section 5.1. TBILL distribution is not shown.

PMVD MCG, SCG, and SCV components are approximately normally distributed. The distribution of all other components are highly skewed and kurtotic. Observe that LCG, LCV, and LT have highly asymmetric distributions which have an exponential character. This profile appears characteristic of variables with relatively weak marginal contribution to model performance but relatively strong contribution to the performance of submodels based on smaller sets of variables.

Figure 2: AMCV component bootstrap distributions

Acronyms are defined in Table 6. AMCV variance components are signed by the sign of the factor beta, see Section 5.1. TBILL is distribution not shown.

Signed AMCV components produce distinctive bimodal bootstrapping distributions. Bimodality is principally a reflection of the averaging method's violation of the proper exclusion criterion. In bootstrap scenarios variables with small statistically insignificant betas, but of opposite sign from the beta based on real data, can have large variance shares.
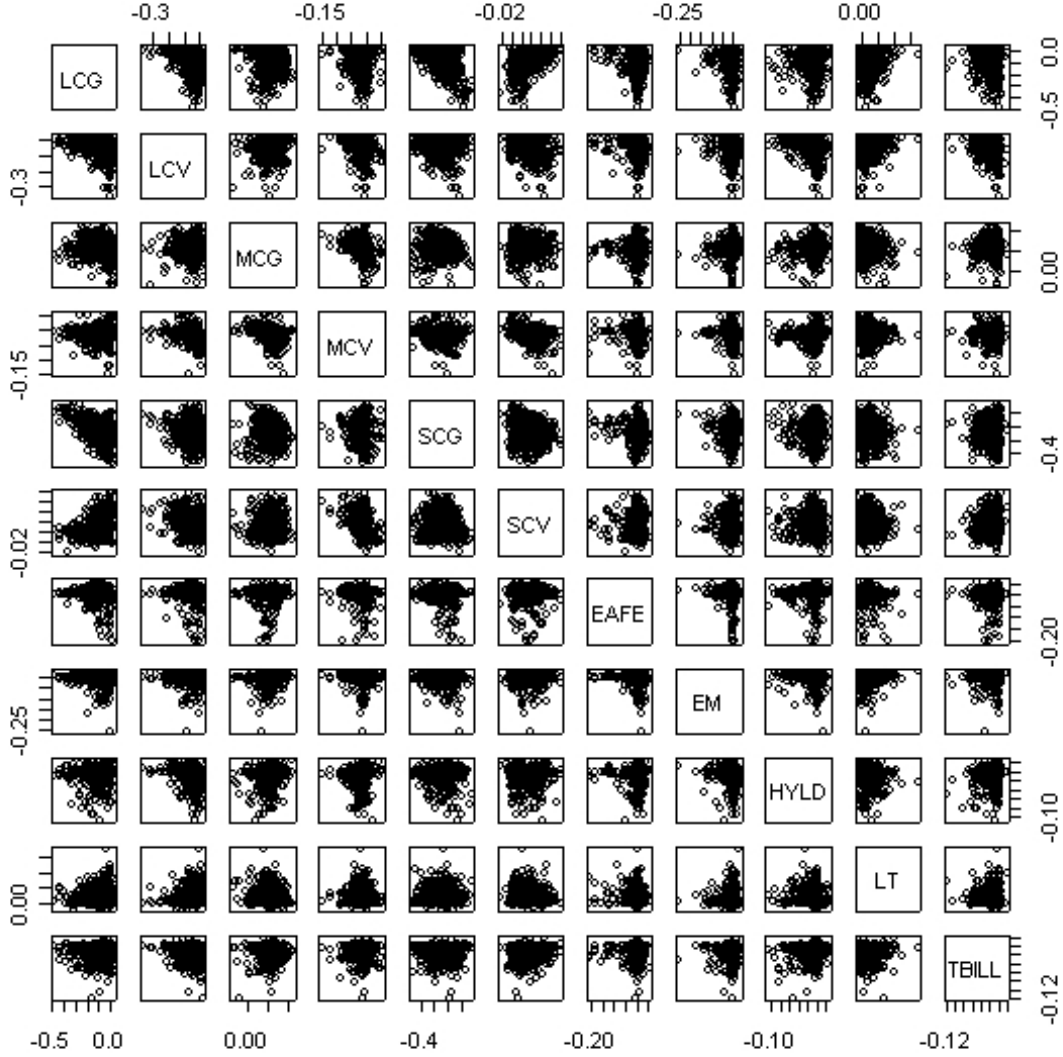
Figure 3: CVD component bootstrap distributions
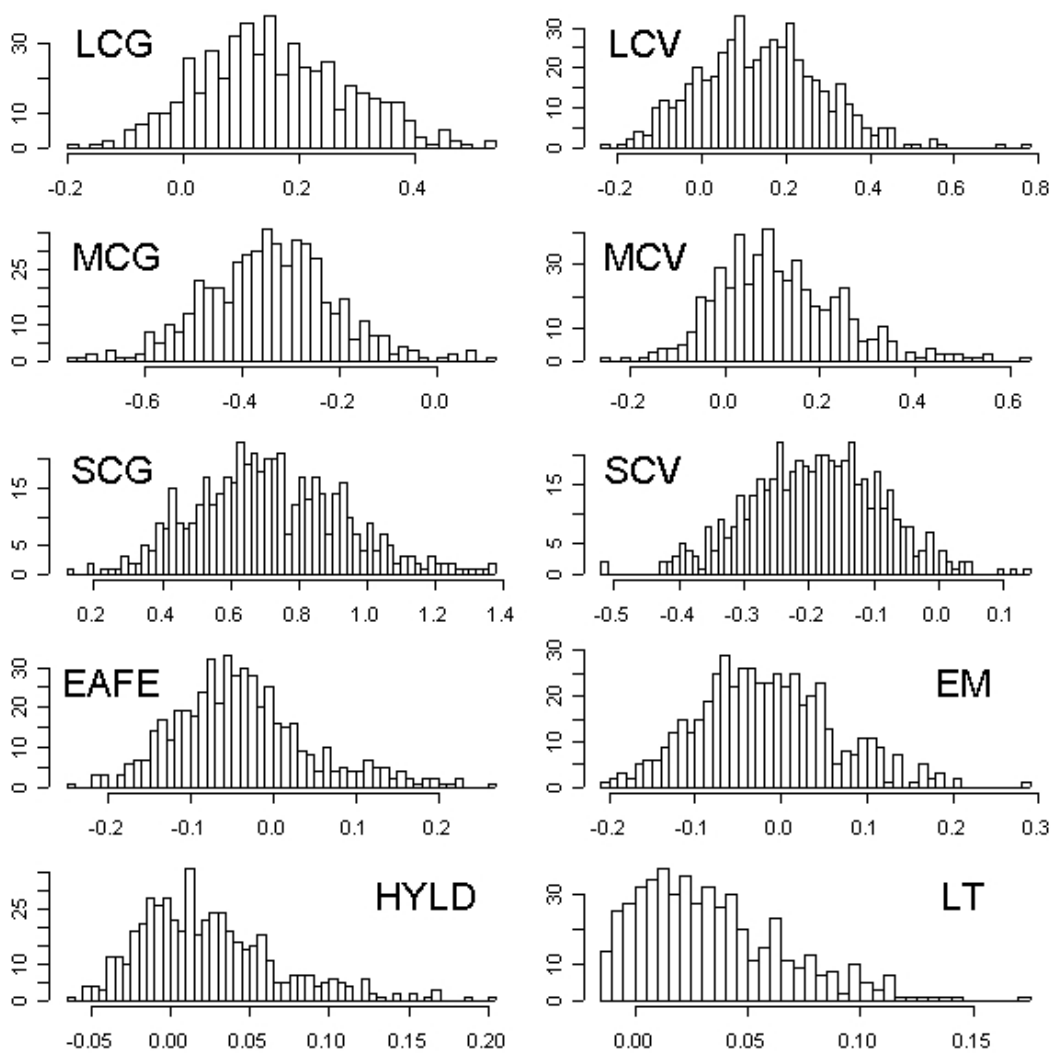
Acronyms are defined in Table 6. TBILL distribution is not shown.

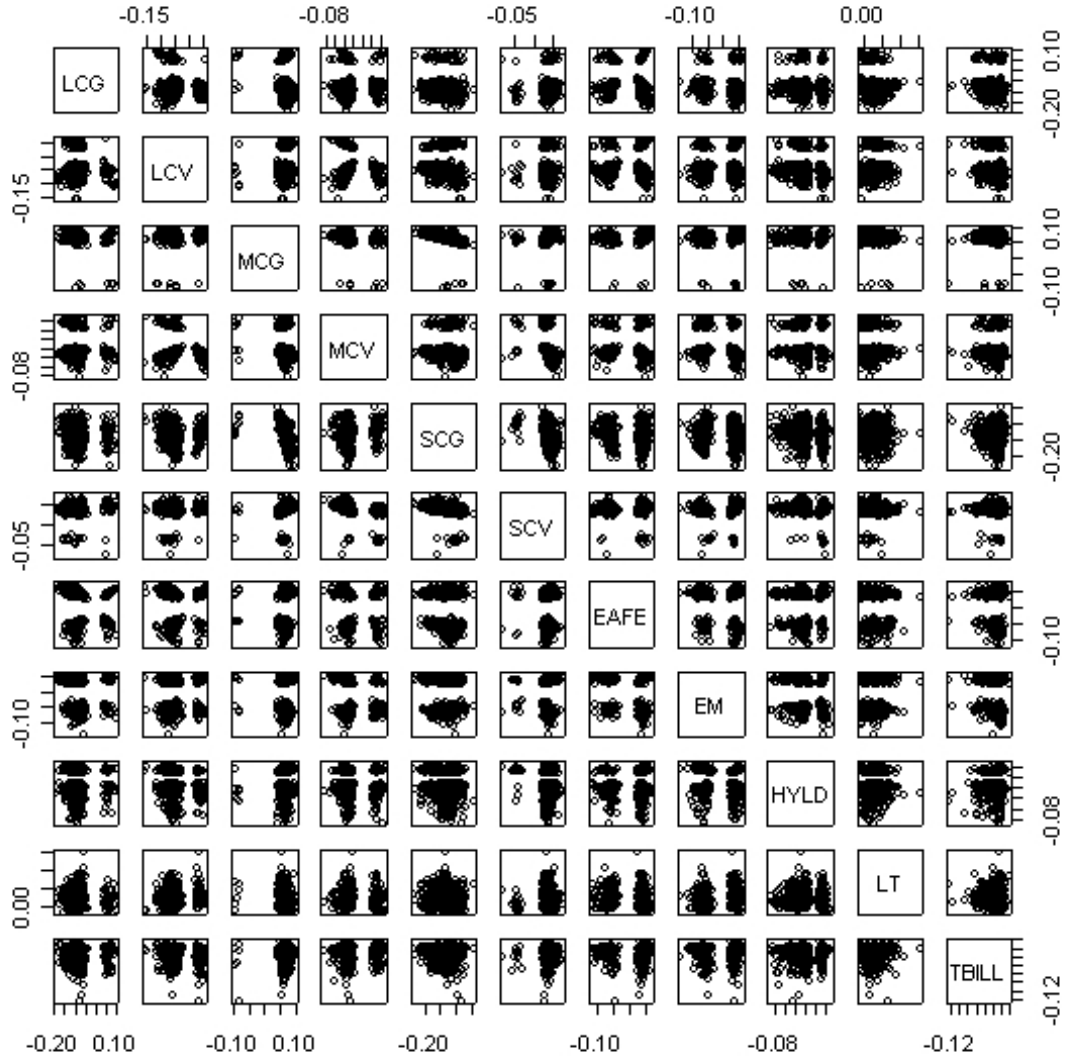**Figure 4: Beta bootstrap bivariate scatter plots**
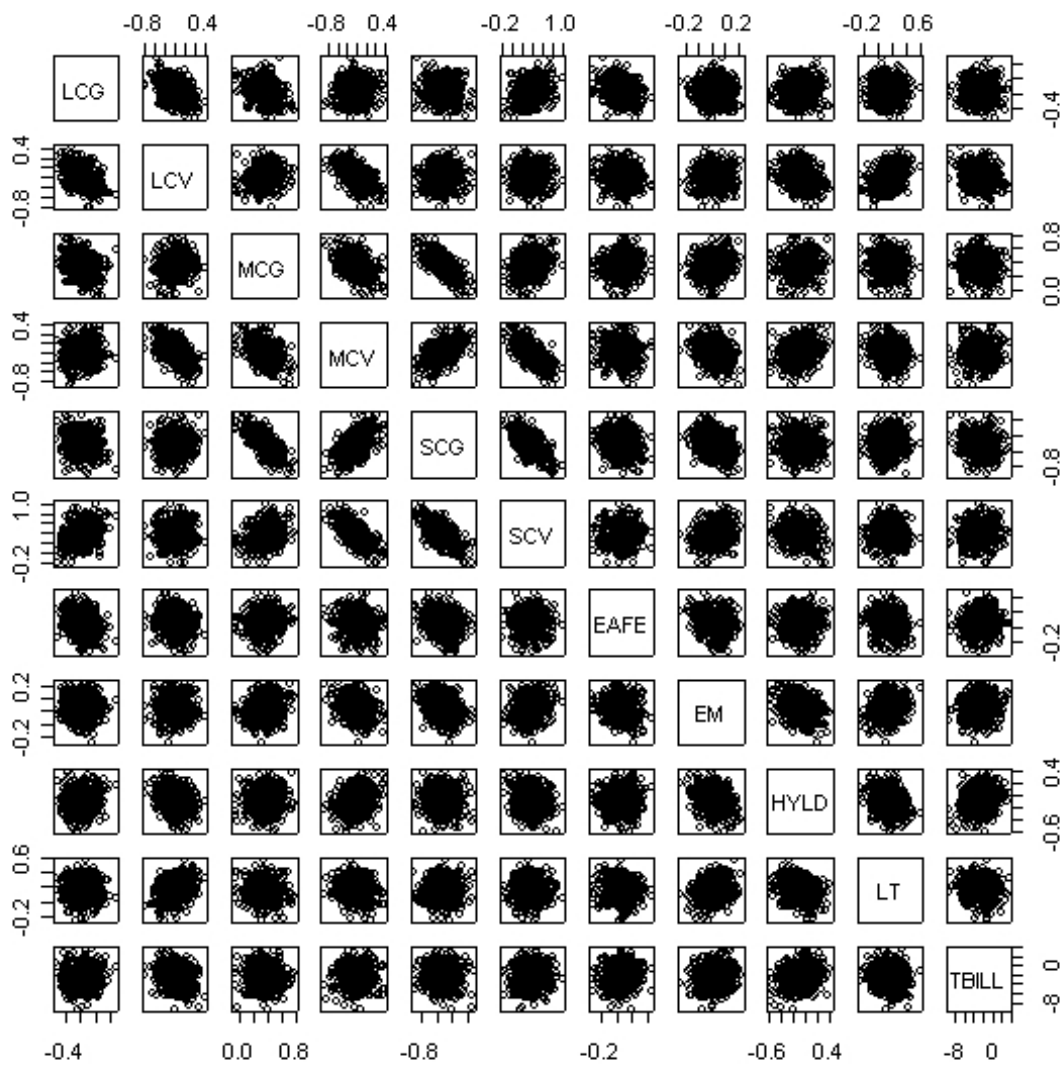
Acronyms are defined in Table 6. Bivariate beta scatters plots are provided for reference. They appear approximately elliptically distributed.

**Figure 5: PMVD bootstrap bivariate scatter plots**

Acronyms are defined in Table 6. PMVD bivariate scatter plots show a distinctive conditional distribution pattern. Let X represent the variable charted on the horizontal axis and Y the variable charted on the vertical axis and define 0-1 scaled variables X^* and Y^* by the following transformation: $X_i^* = (X_i - X_{min})/(X_{max} - X_{min})$. Then LCG has the approximate relationship $LCG^* > (1 - Y^*)$ with all variables except SCV and LT, which are predominantly positively signed, where the relationship is $LCG^* > Y^*$. Both relationships indicate strong conditionality and small values of other components when LCG is large. See Section **??** for more discussion.

Figure 6: AMCV bootstrap bivariate scatter plots

Acronyms are defined in Table 6. Many of the signed AMCV bivariate bootstrap scatter plots have a distinctive clover type of pattern resulting from the bimodal bootstrap beta component distribution shown in Figure 2. This is further evidence that averaging is not a reliable decomposition procedure.

Figure 7: CVD bootstrap bivariate scatter plots
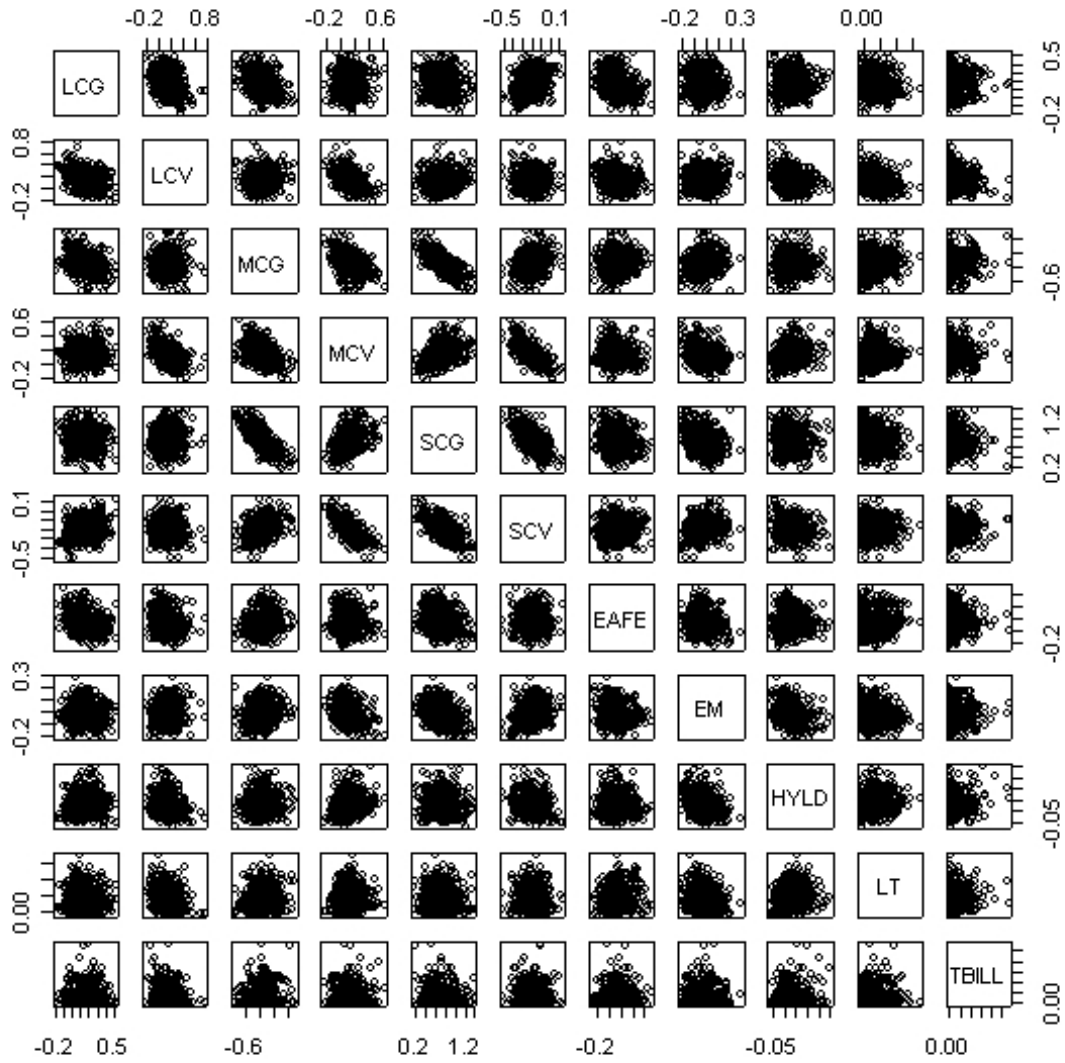
Acronyms are defined in Table 6. The bivariate distribution of CVD bootstrap components is broadly similar to the bivariate distribution of beta coefficients shown in Figure 4.