DIGITALCOMMONS
@WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 7 | Issue 1 Article 4

5-1-2008

On Measuring the Relative Importance of Explanatory Variables in a Logistic Regression

D. Roland Thomas
Carleton University, rthomas@sprott.carleton.ca

PengCheng Zhu
Carelton University

Bruno D. Zumbo
University of British Columbia, bruno.zumbo@ubc.ca

Shantanu Dutta
University of Ontario Institute of Technology

Journal of Modern Applied Statistical Methods
May, 2008, Vol. 7, No.1, 21-38

1538 – 9472/08/$95.00

Regular Articles

On Measuring the Relative Importance of Explanatory Variables in a Logistic Regression

D. Roland Thomas
Carleton University

PengCheng Zhu
Carleton University

Bruno D. Zumbo
University of British Columbia

Shantanu Dutta
St. Francis Xavier University

A search is described for valid methods of assessing the importance of explanatory variables in logistic regression, motivated by earlier work on the relationship between corporate governance variables and the issuance of restricted voting shares (RSF). The methods explored are adaptations of Pratt's (1987) approach for measuring variable importance in simple linear regression, which is based on a special partition of R^2. Pseudo-R^2 measures for logistic regression are briefly reviewed, and two measures are selected which can be partitioned in a manner analogous to that used by Pratt. One of these is ultimately selected for the variable importance analysis of the RSF data based on its small sample stability. Confidence intervals for variable importance are obtained using the bootstrap method, and used to draw conclusions regarding the relative importance of the corporate governance variables.

Key words: Variable Importance, pseudo-R square, corporate governance.

Introduction

This article describes a search for statistical measures to answer the following applied question: How can one determine the relative importance of correlated explanatory variables in a logistic regression? The case that has motivated this study features a sample of firms listed on the Toronto Stock Exchange, some of which issue restricted voting shares, while the remainder do not (Jog, Zhu, & Dutta 2006). Firms that have issued restricted shares to the market will henceforth be referred as restricted share firms (RSF) and the combined dataset featuring both types of firms will be referred to as the RSF dataset.

In the case study, logistic regression is used to quantify the relationship between the issuance of restricted voting shares (issue / do not issue) and three constructed measures of corporate governance, namely dispersion of ownership (DISP), suppression of shareholders interests (SUPP) and board independence (INDEP). The methods that will be constructed to assess the relative importance of these explanatory variables will be quite general and can be applied to a wide range of logistic regression problems. The performance of these methods will be evaluated on a constructed dataset that has known properties, and then applied to the RSF dataset. Practitioners frequently ask how to assess variable importance (Healy, 1990), and when the question relates to explanatory variables in logistic regression, the usual recommendation is to inspect the relative magnitudes of the Wald statistics for individual

D. Roland Thomas is Professor Emeritus at the Sprott School of Business, Carleton University, Ottawa, Canada. Email him at rthomas@sprott.carleton.ca. PengCheng Zhu is an Assistant Professor of Finance at the Eberhardt School of Business, University of the Pacific Stockton campus. Shantanu Dutta is an Assistant Professor of Finance at University of Ontario Institute of Technology. Bruno D. Zumbo is a Professor at the University of British Columbia. E-mail: bruno.zumbo@ubc.ca.

explanatory variables (or their square roots which can be interpreted as large sample z-statistics). The problem with this and related approaches can be easily explained with reference to the governance example. For the explanatory variable DISP, its Wald statistic (or its square root z-statistic) shown in Table 3 is a measure of the contribution of DISP to the logistic regression, over and above the contribution of explanatory variables SUPP and INDEP.

Similarly, the Wald statistic for variable SUPP measures its contribution over and above variables DISP and INDEP. Clearly, it is not appropriate to use these two Wald statistics as measures of the relative contribution of DISP and SUPP because the reference set of variables is different in both cases (SUPP and INDEP in the first case, and DISP and INDEP in the second case). The equivalent problem occurs in linear regression, i.e., the t-statistics (or corresponding p-values) for individual variables are not appropriate for assessing relative importance. Considerable attention has been paid to the problem of variable importance in linear regression, evidenced by the work of Pratt (1987), Kruskall (1987), Budescu (1993), Thomas, Hughes and Zumbo (1998), Azen, Budescu and Reiser (2001), Azen and Budescu (2003), Thomas, Zhu, and Decady (2007), and many others.

Although the interpretational questions that arise in logistic regression are generally similar to those encountered in multiple regression (Hosmer & Lemeshow 2000), no comparable attention has been focused on the question of variable importance in the logistic case. The reason for this lack of attention is more likely due to the greater complexity of the logistic model than to any fundamental difference in interpretational requirements. This complexity is also reflected in measures of fit. For example, while R^2 in multiple regression is a widely accepted and natural measure of model fit, which is easily computed and well understood, analogous measures for logistic regression are not as well known. Though several plausible pseudo-R^2 measures have been proposed and compared for logistic regression (Windmeijer 1995; Mittlbock & Schemper 1996), no one measure has yet been accepted as the standard.

The issue of a pseudo-R^2 for logistic regression is particularly relevant to the subject of this paper. One measure of variable importance in multiple regression that has been extensively discussed in the literature is defined in terms of the portion of “total variance explained” that is assigned to each variable. The rule for partitioning R^2 into its individual components, each representing variable importance, was axiomatically justified by Pratt (1987) and has also been given an easily generalized geometric interpretation by Thomas et al. (1998). Thus, to derive a measure of variable importance for logistic regression, it is natural to seek a pseudo-R^2 measure for logistic regression that can be partitioned in an analogous way. It turns out that not all of the pseudo-R^2 measures proposed to date are suitable for such partitioning. A brief review of the better known measures will be given, one of which (Laitila 1993; McKelvey & Zavoina 1975) can be partitioned in a manner similar to that used by Pratt (1987). An additional pseudo-R^2 measure based on a weighted least squares (WLS) representation of the maximum likelihood estimates (MLE) of the logistic regression parameters is also proposed in this paper. This WLS representation lends itself to partitioning using the geometric approach of Thomas et al. (1998), and so provides an alternative set of importance measures, henceforth referred to in this paper as VI indices.

The article is organized as follows. First, the RSF example and dataset are described, along with results of the basic logistic regression analysis. Also described is a large synthetic dataset with population characteristics designed to mimic the sample data, and which will be used throughout to illustrate the properties of the various methods, and to guide the interpretation of the corporate governance case. Next, Pratt’s (1987) axiomatically derived measure of importance for multiple regression is discussed, which will provide the basis for the various sets of VI indices developed in this paper. Specific attention will be paid to the geometric interpretation given by Thomas et al. (1998). Then, a brief account is given of the pseudo-R^2 measures proposed to date for logistic regression, which (except for the method proposed by Laitila, 1993, and McKelvey & Zavoina1975) cannot be partitioned using either the axiomatic or the geometric approach.

The pseudo-R^2 measure based on the WLS representation of the logistic MLE is then described. VI indices for logistic regression based on the two pseudo-R^2 measures that can be partitioned are then derived, and their particular characteristics are illustrated using the synthetic dataset. Next, these VI indices are used to shed light on the relative importance of the three governance variables, DISP, SUPP and INDEP. This section also describes the bootstrapping techniques used to determine standard errors and confidence intervals for VI indices, which are then used to determine the final variable importance orderings. Finally, an overview and recommendations for future research are given.

Example Datasets

Restricted Shares and Corporate Governance

Restricted shares are a regular feature of the Canadian stock market, and unlike traditional common shares which usually carry one voting right per share, restricted shares have reduced voting rights and in some cases carry no voting rights at all. The issuance of restricted

Table 1. Definition of Study Variables

Variables	Explanation
EXPAY	CEO excess payment
BOARD_SIZE	Size of company board of Directors
P_INS_DIR	Percentage of internal Directors on company board
CEO_CHAIR	If CEO is the Chairman of the board (Yes is 1, No is 0)
DIR_OWN	Percentage of Director ownership
DIR_VOT	Percentage of Director voting rights
COM_OWN	Percentage of combined Director and Block ownership
COM_VOT	Percentage of combined Director and Block voting rights
DIR_OWN_VOT	Ratio of Director voting rights to Director ownership

Table 2. Results of the Factor Analysis

Component Score Coefficient Matrix

	Component			Component Name
	1	2	3	
COM_OWN	0.331	-0.132	-0.159	Dispersion of Ownership and Voting Rights
DIR_OWN	0.311	-0.094	-0.031	
COM_VOT	0.252	0.131	0.061	
DIR_VOT	0.247	0.130	0.143	
DIR_OWN_VOT	-0.071	0.357	0.272	Suppression of Shareholders' Interests
EXPAY	0.006	0.337	-0.145	
BOARD_SIZE	0.017	0.335	-0.150	
P_INS_DIR	0.002	0.000	0.510	Board Independence
CEO_CHAIR	-0.034	0.053	0.517	

Note: Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization shares to the public market reduces the access of non-management shareholders to shares that carry normal voting rights, so that a small number of shareholders (primarily the management group) can effectively control the corporate board. Increasing interest and concern about corporate governance mechanisms in RSFs is now being expressed not only by academic researchers but also by professionals and legislators, particularly in view of the many recent corporate scandals in North America. One of the many objectives of Jog, Zhu and Dutta's (2006) study was to examine the relationship between various corporate governance characteristics and a firm's propensity to issue restricted shares. The final dataset for analysis contained 95 Canadian firms that had restricted shares outstanding on the Toronto Stock Exchange (TSX) between September 1993 and December 2004. A comparison sample was randomly selected from among the TSX companies that had issued no restricted shares during those ten years, providing a combined RSF dataset of 202 firms. A variety of corporate finance and governance variables were collected, as catalogued in Table 1, and a preliminary analysis (not shown) showed the corporate governance variables to be significantly correlated.

A factor analysis and a non-orthogonal "oblimin" rotation was carried out to provide a more succinct and interpretable representation of the variables of Table 1. From Table 2 it can be seen that a useful data summary is provided by three rotated corporate governance factors mentioned in the introduction, namely dispersion of ownership (DISP), suppression of shareholders interests (SUPP) and board independence (INDEP). The estimated correlations between these composites are: (DISP, SUPP) = .06; (DISP, INDEP) = .21 and (SUPP, INDEP) = -.07. Using the SPSS program, scores for each of the corporate governance composite variables were generated using the "regression" method, and saved for subsequent logistic analysis. It should be noted that, in this analysis, no allowance is made for measurement errors arising from the estimation of governance variables that could be regarded as latent. The sampling plan for the Jog et al. (2006) dataset comprises a case-control sample, in which all RSF firms but only a fraction of the non-RSF firms were sampled. However, it is well known (see Hosmer & Lemeshow 2000, p. 178-181) that when the RSF indicator is treated as a binary random variable, consistent regression parameter estimates are obtained for the explanatory variables; only the estimate of the intercept parameter being inconsistent (or biased).

Because Pratt's (1987) variable importance measures do not depend on the intercept parameter, the case-control nature of the sample will not be a problem. Basic results for the logistic regression of the RSF indicator (RSF=1, non-RSF = 0) on the three composite governance variables are shown in Table 3. A Hosmer-Lemeshow goodness-of-fit test suggests that the model does fit the data ($p = 0.31$).

A Synthetic Dataset

A large synthetic dataset containing 50,000 observations was randomly drawn from a population model designed to partially mimic the corporate governance example. The model features three explanatory variables, with regression parameters equal to the MLEs shown in Table 3, and with explanatory variable means and model covariance matrix set equal to the sample means and sample covariance matrix of the three corporate governance variables. Details of the probabilistic structure of the model, which generates samples that are exactly consistent with a logistic regression model, will be given later. The synthetic dataset will be used to compare the various pseudo-R^2 and corresponding sets of VI indices that will be developed, free of the idiosyncrasies typically present in real data. This will facilitate the interpretation of the new measures when they are applied to the RSF data.

Pratt's Measure of Variable Importance for Multiple Linear Regression

The methods used for developing the variable importance measures for logistic regression will all be adaptations of Pratt's (1987) linear regression method which comprises a particular partition of R^2. Pratt's method will be outlined in this section given its central importance to the study. A more detailed summary of Pratt's method is given by Thomas, Zhu, and Decady (2007).

The Axiomatic Approach

Pratt (1987) considered a linear regression equation of the form

$$y = b_0 + b_1 x_1 + \ldots + b_p x_p + u \qquad (1)$$

where the disturbance term u is uncorrelated with $x_1, \ldots, x_p$, and is distributed with mean zero and variance σ^2. The total (standardized) population variance, R_p^2, explained by model (1) can be written as

$$R_p^2 = \sum_j \beta_j \rho_j \qquad (2)$$

where β_j is the usual standardized regression coefficient corresponding to x_j, and ρ_j is the simple correlation between y and x_j. Pratt justified the rule whereby relative importance is equated with variance explained, provided that explained variance attributed to x_j is $\beta_j \rho_j$. This definition of variable importance has been widely used in the applied literature (Green, Carroll and De Sarbo 1978), but as documented by Pratt (1987), it has also been severely criticized. Pratt justified the measure using an axiomatic approach based largely on symmetry and invariance to linear transformation. Subject to his axioms, he showed that his measure is unique. An added bonus is that Pratt's measure allows the importance of a subset of variables to be defined additively, as the sum of their individual importances. Other commonly used measures do not allow for an additive definition.

The Geometric Approach

Thomas et al. (1998) gave a sample interpretation of Pratt's measure based on the geometry of least squares. They considered a sample of N observations fitted to a model of the form (1), so that the observed variables $y, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ comprise vectors in an N-dimensional space. Without loss of generality they assumed that all variables have zero mean, i.e., $\boldsymbol{y}'\boldsymbol{1}_N = \boldsymbol{x}_1'\boldsymbol{1}_N = \ldots = \boldsymbol{x}_p'\boldsymbol{1}_N = 0$, where $\boldsymbol{1}_N$ is an $N \times 1$ vector of ones. In this case, $\hat{\boldsymbol{y}}$, the fitted value of $\boldsymbol{y}$, is the projection of $\boldsymbol{y}$ onto the subspace spanned by the explanatory variables

Table 3

Logistic Regression Results for the Combined RSF Dataset

	$\tilde{\boldsymbol{b}}$	s.e.($\tilde{\boldsymbol{b}}$)	Wald	df	exp($\tilde{\boldsymbol{b}}$)
Intercept	0.196	0.236	0.69	1	1.127
DISP	1.290	0.252	26.30	1	3.633
SUPP	2.495	0.397	39.53	1	12.120
INDEP	0.915	0.238	14.78	1	2.497

Note: Included in this table is the value of $\exp(\tilde{b}_j)$, j = 1, 2, 3, where $\tilde{b}_j$ denotes the MLEs of the logistic regression coefficient for the j'th of the three explanatory variables. The exponential of the j'th regression parameter represents the proportional increase in the odds of a firm being an issuer of restricted voting shares corresponding to an increase of one unit in its score on the j'th explanatory variable, with all other scores held constant. While it is tempting to use these odds ratios as measures of relative importance, it is easily seen that they suffer from precisely the same flaw as do the Wald or z-statistics.

$\boldsymbol{x}_j$, $j = 1, \ldots, p$, and has the representation

$$\hat{\boldsymbol{y}} = \hat{b}_1 \boldsymbol{x}_1 + \ldots + \hat{b}_p \boldsymbol{x}_p , \qquad (3)$$

where the $\hat{b}_j$'s are least squares estimates of the population regression coefficients b_j, j = 1, …, p. Figure 1 illustrates the geometric interpretation of Pratt's importance measures in a two-variable model subspace. In this model subspace, appropriate multiples of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (given by the least squares estimates of the regression coefficients, $\hat{b}_1$ and $\hat{b}_2$, respectively) sum geometrically to $\hat{\boldsymbol{y}}$, the projection of $\boldsymbol{y}$ from its N-dimensional space onto the two-dimensional model subspace. The heavy lines represent the vector projection of each component $\hat{b}_j \boldsymbol{x}_j$ onto $\hat{\boldsymbol{y}}$. Clearly, the orthogonal components sum to zero. Thus it is natural to use the (signed) lengths of the individual projections in the $\hat{\boldsymbol{y}}$ direction (which sum to $\hat{\boldsymbol{y}}$) as measures of the contribution of each $\boldsymbol{x}_j$ to $\hat{\boldsymbol{y}}$, i.e., as measures of variable importance. Thomas et al. (1998) actually defined their VI indices, denoted d_j, as the ratio of the signed length of these projections to the length of $\hat{\boldsymbol{y}}$, and showed that

$$d_j = \hat{\beta}_j \hat{\rho}_j \big/ R^2 , j = 1,\ldots, p, \qquad (4)$$

where hats denote sample estimates, and where R^2 is the usual proportion of sample variance explained. It can be seen that the VI indices defined in equation (4) are sample estimates of Pratt's (1987) measures, normalized by R^2 . Defined in this way, they automatically sum to one. The d_j's are analogous to the

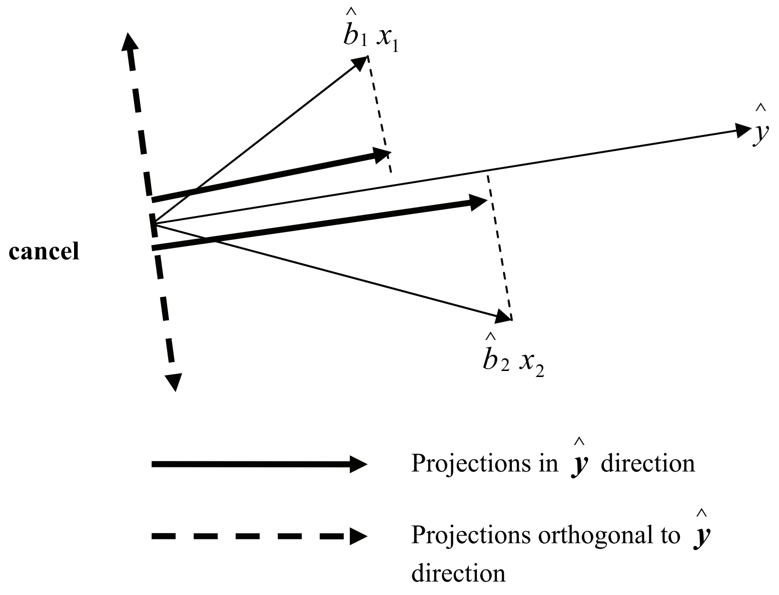

$\hat{b}_1 x_1$
$\hat{y}$
cancel
$\hat{b}_2 x_2$
Projections in $\hat{\boldsymbol{y}}$ direction
Projections orthogonal to $\hat{\boldsymbol{y}}$ direction

Figure 1 Importance Measures as Projections

discriminant ratio coefficients (DRC's) introduced by Thomas (1992) as variable importance measures for descriptive discriminant analysis.

Negative Values of Pratt's VI Indices

Pratt's measure can be negative, a feature that has been criticized and that would appear to detract from its utility as a measure of importance. However, according to Pratt's axiomatic derivation, the importance rule is valid only when the population quantities $\beta_j \rho_j$ are all positive. Thus negativity of any one of these quantities does not signify negative importance, but instead signifies a regression situation which is "too complex for a single measure" (Pratt 1987, p. 245). Thomas et al. (1998) used an extension of the geometric argument to show that negative d_j's of large magnitude can arise from multicollinearity among the predictor variables. They gave an example where a negative VI index of large magnitude (close to one) was reduced to a small positive value by the application of ridge regression (Hoerl & Kennard 1970), suggesting that the original "negative importance" was false. Not all negative importances will be false, however, and the fact must be faced that some regression modeling situations are so complex that there is no single measure of variable importance that satisfies Pratt's axioms. For multiple linear regression, Thomas, Zhu and Decady (2007) have developed simultaneous confidence interval procedures that can be used to identify such cases.

Pratt's axiomatic derivation provides a theoretical foundation for his measure in the case of multiple regression, but it is not necessarily easy to generalize his method to other analyses. The benefit of the geometric interpretation is that it is sometimes easier to apply to other modeling techniques than is the axiomatic approach, as exemplified by Thomas's (1992) introduction of DRC's in discriminant analysis. It will be shown in Section 5 that both the axiomatic and geometrical interpretations of Pratt's method can be extended to the case of logistic regression.

R^2 Measures for Logistic Regression

The Model Setup

The logistic model of interest can be expressed as

$$\log[\pi_i/(1-\pi_i)] = \boldsymbol{x}_i'\boldsymbol{b}, \quad i = 1, \ldots, N, \tag{5}$$

where $\pi_i = P(y_i = 1|\boldsymbol{x}_i)$, and where in this logistic case, y_i, $i = 1, \ldots, N$ are independent binary random variables, $\boldsymbol{x}_i$, $i = 1, \ldots, N$ are $(p+1)$-vectors of observed explanatory variable scores (with first element equal to one) for the i'th individual, and $\boldsymbol{b}$ is a $(p+1)$-vector of regression coefficients (with first element b_0 corresponding to the intercept).

The reader is warned not to confuse the $\boldsymbol{x}_i$, $i = 1, \ldots, N$, notation used in equation (5), which refers to N sample realizations of a $(p+1)$-vector, with the notation $\boldsymbol{x}_j$, $j = 1, \ldots, p$, used in the previous section, which referred to p realizations of an N-vector. The indexing will always be clearly specified to avoid confusion. Also, no notational distinction is made in the paper between a random variable and its realization; the distinction will be clear from the context. In equation (5) it will be assumed that at least one of the predictors will be measured on a continuous scale, so that none of the covariate patterns will be repeated. This is the sparse case in which the Pearson chi-square and the deviance (discussed, for example, by McCullagh and Nelder 1989) do not exhibit their "usual" asymptotic chi-squared distributions, and for which appropriate goodness-of-fit measures are still an issue. The aim of this section is to identify, for the above setup, measures of fit of the R^2 type that can be partitioned to yield VI indices for logistic regression. Some of the relevant R^2 measures proposed to date will be briefly reviewed.

Pseudo-$\boldsymbol{R}^2$ Measures for Logistic Regression

In a review of pseudo-R^2 measures for binary choice models, Windmeijer (1995) reviewed several categories of measures of fit, the first of which is usually attributed to Efron (1978) though it has been considered by a number of authors. It has the form

$$R_E^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \tilde{\pi}_i)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}, \qquad (6)$$

where $\bar{y}$ is the sample mean of the binary y_i's and the $\tilde{\pi}_i$'s in this case denote maximum likelihood estimates (MLEs) of the π_i's. In fact, any consistent estimates of the π_i's will suffice. Mittlbock and Schemper (1996) favored this measure over many of its competitors. However, Cameron and Windmeijer (1996) noted that the lower bound for Efron's measure is not in general equal to zero, and may in some cases be negative. For this reason, and because it cannot readily be partitioned to identify the contribution of individual predictor variables, Efron's measure will not be considered further.

The second category consists of measures based on the loglikelihood corresponding to model (5), namely

$$\log L(\boldsymbol{b}) = \sum_i [y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i)]. \qquad (7)$$

McFadden's (1974) measure has the form

$$R_{MF}^2 = 1 - \log\ L(\tilde{\boldsymbol{b}}) / \ \log L_0, \qquad (8)$$

where $L(\tilde{\boldsymbol{b}})$ denotes the likelihood evaluated at the maximum likelihood estimate $\tilde{\boldsymbol{b}}$, and L_0 denotes the likelihood for the model containing only an intercept term. When there are no repeated predictor patterns, R_{MF}^2 lies in the interval [0, 1]. Otherwise, its upper limit is less than one, in which case the statistic can be adjusted to recover the appropriate limits (Hosmer & Lemeshow 2000, pp. 164). McFadden's measure possesses several attractive features. It is related to the asymptotic chi-squared test that a subset of the model parameters are zero, and it also has an information theoretic interpretation (see Windmeijer, 1995). Unfortunately, it cannot be partitioned into individual importances, either by means of the linear geometric interpretation described earlier or by any other means known to the authors. A related measure due to Cox and Snell (1989) is also based on the likelihood ratio, and has the form

$$R_{CS}^2 = 1 - [L_0 / L(\tilde{\boldsymbol{b}})]^{2/N}. \qquad (9)$$

This measure does not attain an upper limit of one when the model fits perfectly, and it was suggested by Cragg and Uhler (1970) that it should be scaled to give the required upper bound. Nagelkerke (1991) advocated the same scaling and showed that the scaled measure possesses theoretically attractive features. However, Mittlbock, and Schemper (1996) criticized this scaling as cosmetic, noting that there is no theoretical reason why such a scaling should be appropriate at intermediate values of the statistic. As with the McFadden measure, there appears to be no way to partition R_{CS}^2 or its scaled counterpart to account for contributions of individual variables.

A third category of R^2 measures is based on the interpretation of logistic regression (and other binary choice models) as a linear regression of predictors on an unobservable continuously distributed random variable y_i^*, where the observed binary variable y_i takes the value 1 for $y_i^* \geq 1$, and the value 0 for $y_i^* < 1$. The linear model is specified as

$$y_i^* = \boldsymbol{x}_i^{'}\boldsymbol{b} + \varepsilon_i, \qquad i = 1, \ldots N \qquad (10)$$

where the ε_i are independently distributed logistic variables with mean zero and variance $\pi^2/3$, with $\boldsymbol{x}_i'$ and $\boldsymbol{b}$ defined as in equation (10). Had the response variable $\boldsymbol{y}^* = (y_1^*, \ldots, y_N^*)'$ been observed, then standard OLS parameter and residual estimation could be used resulting in a measure R^{*2}. Although $\boldsymbol{y}^*$ is not observable, R^{*2} can be replaced by a measure proposed by McKelvey and Zavoina (1975) and

Laitila (1993), henceforth referred to as MZL, namely

$$R^2_{MZL} = \hat{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{QX}\hat{\boldsymbol{b}}' \big/ (\hat{\boldsymbol{b}}'\,\boldsymbol{X}'\boldsymbol{QX}\hat{\boldsymbol{b}} + N\pi^2/3). \qquad (11)$$

Here $\hat{\boldsymbol{b}}$ is any consistent estimator of $\boldsymbol{b}$, $\boldsymbol{X}$ is an $N \times (p+1)$ matrix having rows $\boldsymbol{x}_i'$, $i = 1, \ldots N$, and $\boldsymbol{Q}$ is the $N \times N$ projection matrix given by $\boldsymbol{Q} = \boldsymbol{I}_N - \boldsymbol{1}_N\boldsymbol{1}_N'/N$. Laitila (1993) gave a more general version of (11) applicable also to limited dependent variable models, in which the error variance term was consistently estimated. The key property of R^2_{MZL} (and its more general versions) is that the difference between it and R^{*2} vanishes with increasing sample size, i.e., it is asymptotically zero in probability. It will be shown in the next section that Pratt's approach can be applied to partition R^2_{MZL} to yield a set of VI indices. It will also be shown that even though the original R^{*2} itself is unobservable, it nevertheless provides the basis for deriving an alternative set of normalized VI indices that are asymptotically close to those derived from R^2_{MZL}.

An R^2 Measure Based on Weighted Least Squares

It was noted by Pregibon (1981) that the maximum likelihood estimator $\tilde{\boldsymbol{b}}$ of $\boldsymbol{b}$ can be represented in terms of the weighted least squares regression of a vector of pseudo-values $\boldsymbol{z}$ on $\boldsymbol{X}$, given by

$$\tilde{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{VX})^{-1}\boldsymbol{X}'\boldsymbol{Vz}, \qquad (12)$$

where $\boldsymbol{z} = \boldsymbol{X}\tilde{\boldsymbol{b}} + \boldsymbol{V}^{-1}\boldsymbol{r}$, $\boldsymbol{r} = (\boldsymbol{y} - \tilde{\boldsymbol{\pi}})$, $\boldsymbol{y}$ is the $N \times 1$ vector of binary observations, $\tilde{\boldsymbol{\pi}}$ is the $N \times 1$ vector of estimated probabilities corresponding to the maximum likelihood estimate $\tilde{\boldsymbol{b}}$, and $\boldsymbol{V}$ is the $N \times N$ diagonal weight matrix having elements $\tilde{\pi}_i\left(1-\tilde{\pi}_i\right)$, *i* = 1, …, *N*. Pregibon (1981) exploited equation (12) to extend the diagnostic techniques of linear regression to logistic regression, and Nordberg (1981) and Hosmer, Jovanovic, and Lemeshow (1989) used it to apply the techniques of all subsets variable selection to logistic regression. In this section the representation (12) will be used to develop a pseudo-R^2 measure for logistic regression.

It will be more convenient to represent equation (12) as the OLS regression of $\boldsymbol{\omega} = \boldsymbol{V}^{1/2}\boldsymbol{z}$ on $\boldsymbol{V}^{1/2}\boldsymbol{X}$, with fitted values $\hat{\boldsymbol{\omega}} = \boldsymbol{V}^{1/2}\boldsymbol{X}\tilde{\boldsymbol{b}}$ and residuals given by

$$\boldsymbol{\omega} - \hat{\boldsymbol{\omega}} = \boldsymbol{V}^{1/2}(\boldsymbol{z} - \boldsymbol{X}\tilde{\boldsymbol{b}}) = \boldsymbol{V}^{-1/2}\boldsymbol{r} \qquad (13)$$

The residual sum of squares from this pseudo-regression is

$$SS_E = (\boldsymbol{\omega} - \hat{\boldsymbol{\omega}})'(\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}) = \boldsymbol{r}'\boldsymbol{V}^{-1}\boldsymbol{r}$$
$$= \sum_{i=1}^{N} \frac{(y_i - \tilde{\pi}_i)^2}{\tilde{\pi}_i(1 - \tilde{\pi}_i)} = \chi^2, \qquad (14)$$

the familiar Pearson "chi-squared" statistic. Alternatively,

$$SS_E = \boldsymbol{\omega}'\boldsymbol{M\omega} = \boldsymbol{z}'\boldsymbol{V}^{1/2}\boldsymbol{MV}^{1/2}\boldsymbol{z} \quad (15)$$

where $\boldsymbol{M}$ is a $N \times N$ projection matrix of rank *N-p-1* given by

$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{V}^{1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{VX})^{-1}\boldsymbol{X}'\boldsymbol{V}^{1/2}. \ (16)$$

This projection matrix, derived from the weighted least squares representation of the maximum likelihood estimate, was used by Pregibon (1981) in his development of logistic regression diagnostics. The OLS version of the maximum likelihood identity also yields a regression sum of squares, given by

$$SS_R = \hat{\boldsymbol{\omega}}'\hat{\boldsymbol{\omega}} - \hat{\boldsymbol{\omega}}'\boldsymbol{V}^{1/2}\boldsymbol{1}\left(\boldsymbol{1}'\boldsymbol{V1}\right)^{-1}\boldsymbol{1}'\boldsymbol{V}^{1/2}\hat{\boldsymbol{\omega}}$$
$$= \tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{V}^{1/2}\boldsymbol{Q}_V\boldsymbol{V}^{1/2}\boldsymbol{X}\tilde{\boldsymbol{b}}, \qquad (17)$$

where $\boldsymbol{Q}_V = \boldsymbol{I} - \boldsymbol{V}^{1/2}\boldsymbol{1}\left(\boldsymbol{1}'\boldsymbol{V1}\right)^{-1}\boldsymbol{1}'\boldsymbol{V}^{1/2}$. Note that equation (17) comprises a weighted version of the numerator of R^2_{MZL} given in equation

(11). Equations (14) and (17) immediately lead to an R^2 measure given by

$$R^2_{WLS} = SS_R /(SS_R + \chi^2) \qquad (18)$$

$$= 1 - \chi^2 /(SS_R + \chi^2). \qquad (19)$$

The geometric interpretation of Pratt's measures will be used in the next section to partition R^2_{WLS} and yield a set of normalized VI indices.

A Numerical Comparison of the Pseudo-R^2 Measures

Pseudo-R^2 measures for the synthetic and RSF datasets are displayed in Table 4. Results for the synthetic data, shown in the leftmost column, can be regarded as population values essentially free of sampling error. Values of Efron's R^2_E, McFadden's R^2_{MF} and Cox and Snell's R^2_{CS} are shown for reference only, as they cannot be partitioned and thus do not provide the basis for the development of VI indices. As is typical of such pseudo-R^2 measures, they vary considerably in magnitude (Mittlbrock & Schemper 1996) for both datasets.

Of all the measures in Table 4, the largest value is recorded by R^2_{MZL} (McKelvey & Zavoina 1975; Laitila 1993), which is not surprising because it is designed to measure the explained variation in the continuous latent variable $\boldsymbol{y}^*$, rather than the variation in the observed vector of binary variables $\boldsymbol{y}$. On the other hand, the new weighted least squares measure R^2_{WLS} records the smallest pseudo-R^2 value of all, for both synthetic and RSF data. It is interesting to note that Mittlbrock and Schemper (1996) argued against using the weighted least squares representation of the MLE to construct a pseudo-R^2 because of the potentially distorting effect of the weights. Generally speaking, the trends exhibited in Table 4 for pseudo-R^2 values are similar for both the RSF dataset and the synthetic dataset, which represents population values. The sample values obtained for the RSF dataset can therefore be validly used for interpretational purposes, even though the sample size is not large.

Variable Importance Indices for Logistic Regression Measures of Importance Based on $\boldsymbol{R}^2_{MZL}$ and R^{*2}

The continuous model (10) satisfies the assumptions of Pratt's axiomatic approach to variable importance for linear models, the only difference being that the dependent variable $y*$ is not observable. Thus the VI indices of equation (4) can be applied provided only that consistent estimates of β_j, ρ_j and R^2 can be obtained. An estimate of R^2 is given by R^2_{MZL}, as described in the previous section, and a consistent estimate of β_j is given by

$$\hat{\beta}^{MZL}_j = \tilde{b}_j \hat{\sigma}_j / \sqrt{\tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{QX}\tilde{\boldsymbol{b}} + N\pi^2 / 3}, \qquad (20)$$

(see equation 11) where $\tilde{b}_j$ is the (known) MLE of the regression coefficient for the j'th predictor variable x_{ij}, $i = 1, \ldots, N$, and where $\hat{\sigma}^2_j = [\boldsymbol{X}'\boldsymbol{QX}]_{\boldsymbol{jj}}$ is its sample variance.

The correlation ρ_j between $\boldsymbol{y}^*$ and each observed predictor $\boldsymbol{x}_j$, $j = 1, \ldots, p$ can be estimated as a polyserial correlation, $\hat{\rho}^{PS}_j$ (Drasgow 1986), inferred using only the observed binary responses y_i, $i = 1, \ldots, N$ and the observed predictors. These estimates together yield the set of VI indices

$$d^{MZL}_j = \hat{\beta}^{MZL}_j \hat{\rho}^{PS}_j \Big/ R^2_{MZL}, \quad j = 1, \ldots, p. \qquad (21)$$

Note, however, that this application of polyserial coefficients invokes an assumption of joint multivariate normality of y^*_i and $\boldsymbol{x}_i$ which is not required in the development of R^{*2} or R^2_{MZL}. Further, since $\boldsymbol{\varepsilon_i}$ in equation (10) is assigned a logistic distribution, y^*_i itself will not be multivariate normal. Calculations based on the synthetic dataset yield $R^2_{MZL} = 0.741$, and $\sum_j \overset{\wedge}{\beta}{}^{MZL}_j \; \overset{\wedge}{\rho}{}^{PC}_j = 0.747$, indicating that for normal $\boldsymbol{x_i}$'s and logistic $\boldsymbol{\varepsilon_i}$, estimates of the polyserial correlations are robust to this violation of joint normality when the predictors themselves are normal. (The corresponding comparison for the RSF data yields 0.737 versus 0.758). Despite this robustness, it is nevertheless worth seeking normalized VI indices that do not rely on polyserial correlation estimates.

An alternative expression for VI indices can be obtained by applying the derivation of Thomas, Hughes and Zumbo (1998) with $\boldsymbol{y}^*$ treated as known. This yields

$$d^*_j = [\, \boldsymbol{y}'^* \boldsymbol{Qx}_j \, \hat{b}_j \,/\, N\,] \,/\, R^{*2} = [\, \hat{\boldsymbol{b}}' \boldsymbol{X}' \boldsymbol{Qx}_j \, \hat{b}_j \,/\, N\,] \,/\, R^{*2} \qquad (22)$$

where $\hat{\boldsymbol{b}}'$ and $\hat{b}_j$ represent OLS regression parameter estimates. Thus knowledge of $\boldsymbol{y}^*$ is not needed to define VI indices; consistent estimates of the population values of d^*_j can be obtained by replacing $\hat{\boldsymbol{b}}$ by the MLE $\tilde{\boldsymbol{b}}$, and R^{*2} by R^2_{MZL}. As a result of these replacements, the sum of the d^*_j's will sum to one asymptotically, without the slight approximation inherent in the method that relies on the polyserial coefficient. Furthermore, normalized VI indices that sum identically to one can be defined as

$$d^*_j(N) = \frac{\tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{Qx}_j\tilde{b}_j}{\sum_j \tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{Qx}_j\tilde{b}_j} = \frac{\tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{Qx}_j\tilde{b}_j}{\tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{QX}\tilde{\boldsymbol{b}}}, \qquad (23)$$

where the denominator of equation (23), divided by N, is asymptotically equivalent to R^{*2} and R^2_{MZL}. Equation (23) represents the most convenient version of a VI index based on the linear representation (10). Values of both d^{MZL}_j and $d^*_j(N)$ for the synthetic dataset are displayed in Table 5.

Measures of Importance Based on $\boldsymbol{R}^2_{WLS}$

The assumptions underlying Pratt's axiomatic approach do not apply to the WLS representation of the MLE given in equation (12). However, the measure of fit R^2_{WLS} can be partitioned by applying the geometric approach of Thomas et al. (1998) to the pseudo-regression formulation of Section 4, i.e., by an appropriate interpretation of equation (4). Let $\tilde{\beta}_j$ represent the standardized logistic regression coefficient corresponding to the *j*th predictor, *j* = 1, …, *p*, given by

$$\tilde{\beta}_j = b_j (x_j' V^{1/2} Q_V V^{1/2} x_j)^{1/2} \Big/ (z' V^{1/2} Q_V V^{1/2} z)^{1/2} \qquad (24)$$

where $\tilde{b}_j$ is the maximum likelihood estimate of the *j*th logistic regression coefficient *b*

$$\tilde{\rho}_j = z' V^{1/2} Q_V V^{1/2} x_j \Bigg/ \left\{ \left(z' V^{1/2} Q_V V^{1/2} z \right)^{1/2} \left(x_j' V^{1/2} Q_V V^{1/2} x_j \right)^{1/2} \right\} \qquad (25)$$

Then the required VI indices for the *j'*th predictor variable (i.e. the *j'* th partition of R^2_{WLS}) are obtained from equation (4) as

$$d^{WLS}_j = \tilde{\beta}_j \, \tilde{\rho}_j \Big/ R^2_{WLS} \,. \qquad (26)$$

Table 4. Pseudo-R^2 Measures for the Synthetic and Restricted Share Firms Datasets

Pseudo $\boldsymbol{R^2}$ Measures		Synthetic Data (N=50,000)	RSF Data (N = 202)
R^2_{MZL}	(equation 11)	0.741	0.737
R^2_{WLS}	(equation 18)	0.193	0.226
R^2_E	(equation 6)	0.549	0.483
R^2_{MF}	(equation 8)	0.487	0.507
R^2_{CS}	(equation 9)	0.491	0.504

Table 5. Variable Importance Indices for the Synthetic and RSF Datasets

VI Indices		Synthetic Data (N=50,000)			RSF Data (N = 263)		
		DISP	SUPP	INDEP	DISP	SUPP	INDEP
d_j^{MZL}	(equation 21)	.226	.683	.096	.297	.619	.113
$d_j^*(N)$	(equation 23)	.224	.680	.096	.227	.674	.099
d_j^{WLS}	(equation 26)	.224	.682	.094	.374	.499	.127
$d_j^{WLS}(N)$	(equation 27)	.224	.682	.094	.374	.499	.127

In practice, with dependent variable $V^{1/2}z$ and predictor variables $V^{1/2}X$, the quantities $\tilde{\beta}_j$, j = 1, …, p, and R^2_{WLS} can be obtained from the output of standard multiple regression programs as the standardized regression coefficient (the "beta" weight in SPSS, for example) and the standard R^2 measure, respectively. Similarly, the correlation $\tilde{\rho}_j$ corresponds to the standardized regression coefficient in the simple linear regression of $\boldsymbol{V}^{1/2}\boldsymbol{z}$ on $\boldsymbol{V}^{1/2}\boldsymbol{X}$.

An algebraically equivalent representation of d_j^{WLS} can be derived in a manner similar to that used to derive equation (22), by applying regression identities to the WLS representation of the MLE $\tilde{\boldsymbol{b}}$ defined by equation (12). This leads to the expression

$$d_j^{WLS}(N) = \frac{\tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{V}^{1/2}\boldsymbol{Q}_v\boldsymbol{V}^{1/2}\boldsymbol{x}_j\tilde{b}_j}{\tilde{\boldsymbol{b}}'\boldsymbol{X}'\boldsymbol{V}^{1/2}\boldsymbol{Q}_v\boldsymbol{V}^{1/2}\boldsymbol{X}\tilde{\boldsymbol{b}}}, \tag{27}$$

which yields a weighted least squares analogue of equation (23).

A Numerical Comparison of the Competing VI Indices

Values of the variable importance indices described in the previous section are shown in Table 5 for the three corporate governance variables. The third and fourth rows of the table simply illustrate the fact that equations (26) and (27) are algebraically equivalent, with VI indices that sum to one. These alternative forms will be referred to in what follows as d_j^{WLS}.

It can be seen from the first row of Table 5 that the VI indices d_j^{MZL} do not sum to one exactly, as was explained in the text. For the synthetic dataset they sum to 1.008 and for the RSF dataset they sum to 1.028, both representing only minor discrepancies. The indices $d_j^*(N)$ do sum to one by virtue of their construction, and will be used henceforth in preference to d_j^{MZL}. Thus the important conclusions to be drawn from Table 5 relate to the two index sets $d_j^*(N)$ and d_j^{WLS}.

For the former, it can be seen that individual indices for the three independent variables are very similar for both the synthetic (population) dataset and the RSF dataset. This suggests that the VI indices $d_j^*(N)$ perform well for moderate size samples, a conclusion that should be explored in greater detail in a more extensive simulation study. It can also be seen from Table 5 that both sets of indices, $d_j^*(N)$ and d_j^{WLS}, exhibit very similar results for the large sample synthetic dataset. However, for the moderately sized RSF dataset, the d_j^{WLS} indices differ noticeably from these large sample values, suggesting that the VI indices d_j^{WLS} might be less robust to small and medium sample sizes than the indices $d_j^*(N)$.

It was noted earlier that Mittlbock and Schemper (1996) recommended against using the weighted least squares representation of the logistic regression MLE because of the potentially distorting effect of the weights. While these weights appear to have little impact on either set of VI indices for the large sample synthetic dataset, their effect may be more severe for smaller sample sizes. For this reason, the following analysis of variable importance in the RSF dataset will be based entirely on the $d_j^*(N)$ indices.

An Analysis of Variable Importance for the RSF Dataset

Point Estimates of Importance

The point estimates of the VI indices $d_j^*(N)$ suggest that SUPP (suppression of shareholders' interests) is the most important governance variable for differentiating between restricted share firms and non-restricted shares firms, and that INDEP (board independence) is the least important, with the effect of DISP (dispersion of ownership) being intermediate. However, to decide if these differences between point estimates translate into real (population) differences in variable importance, standard errors and confidence intervals for each individual index must be estimated. Thomas, Zhu, and Decady (2007) provided large sample formulas for the standard errors of normalized Pratt indices for the linear regression case, but it is not practical to extend their analysis to the logistic regression case. However, because the VI indices proposed in this paper are smooth functions of means, variance and covariances, standard errors can be obtained using the bootstrap resampling methodology, as described in the following section.

Standard Errors and Confidence Intervals for the VI Indices

A standard non-parametric bootstrap (Efron and Tibshirani 1993) was used to estimate the standard errors and corresponding confidence intervals for the indices $d_j^*(N)$. The resampling procedure consisted of 1000 independent bootstrap samples of 200 observations (each taken with replacement from the original RSF sample). From the 1000 bootstrap samples, 1000 replications of the logistic parameter estimates and VI indices were then calculated, allowing for the computation of bootstrap standard errors, as well as a visual depiction of the bootstrap distribution. All computations were carried out using the bootstrap facilities of the R language (Canty and Ripley 2006). Histograms of the bootstrap samples for the VI indices $d_j^*(N)$ are shown in Figure 2, and corresponding bootstrap standard errors are shown in Table 6.

Figure 2

Bootstrap Histograms of The VI Indices $d_j^*(N)$

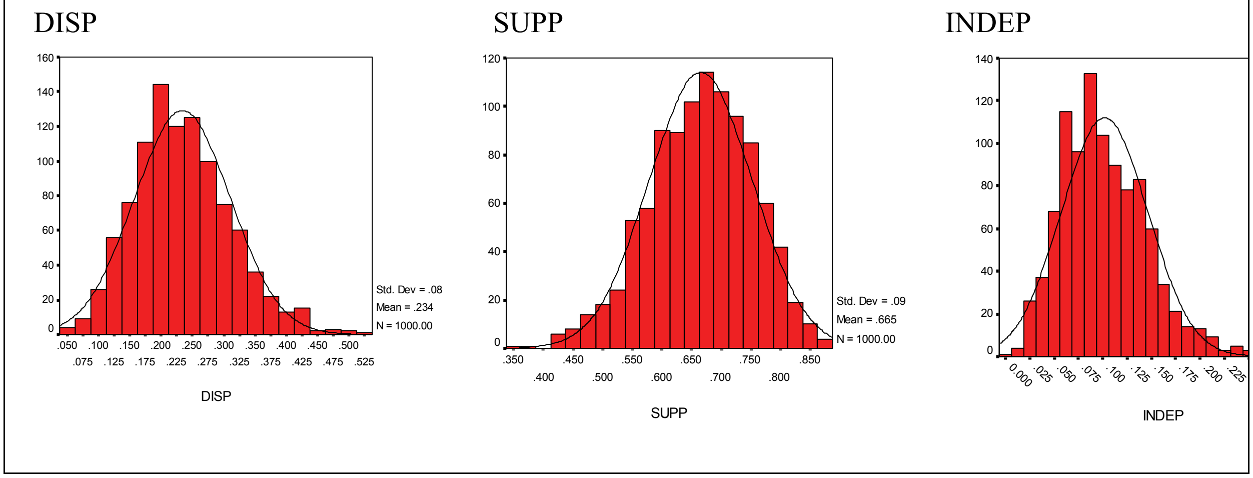
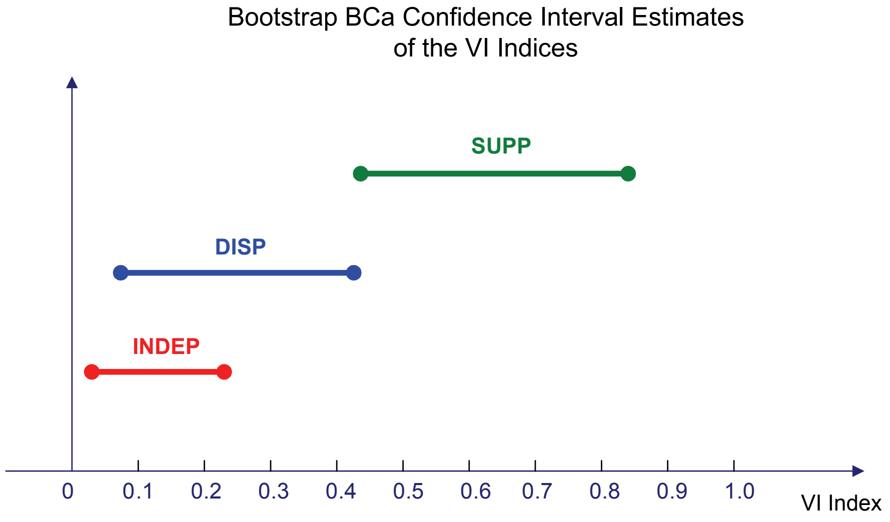


Table 6

Bootstrap Standard Errors and BCa Confidence Intervals for VI Indices $d_j^*(N)$

Variables	Point Estimates	Standard Errors	Individual 95% CIs	Simultaneous 95% CIs
DISP	.227	.077	(.095, .400)	(.072, .434)
SUPP	.674	.087	(.490, .831)	(.440, .841)
INDEP	.099	.044	(.032, .214)	(.018, .232)

Figure 3



Large sample confidence intervals are often computed simply as a point estimate plus and minus twice the standard deviation of the statistic in question. However, in cases where the sampling distribution still retains some non-normality, such confidence intervals tend to provide poor coverage. Numerous alternatives based on the bootstrap have been described in the literature (Efron and Tibshirani 1993; Davison and Hinkley 1997), and it has been shown that the Bias Corrected and Accelerated (BCa) interval has superior coverage properties (Platt, Hanley and Yang 2000). A major advantage of the BCa interval is its transformation-respecting property, i.e., the method effectively selects the best (most normal) scale and then transforms the interval back to the original scale of interest (Efron 1987). Individual BCa 95% confidence intervals (CIs) for the VI indices $d_j^*(N)$ are shown in Table 6 along with the point estimates and standard errors.

Individual confidence intervals are appropriate if the VI index of a specific variable is of prior interest. If the interest results from the analysis itself, i.e., if the VI index of a particular variable is the largest, which implies a comparison with all other variables, then simultaneous confidence intervals should be used (Thomas et al. 2007). As shown by the latter authors, simultaneous confidence intervals can be obtained using the Bonferroni adjustment which, for the RSF case featuring three explanatory variables, implies constructing confidence intervals each at a nominal alpha level of 100(1 - .05/3)%. These also are shown in Table 6.

From Table 6 it can be seen that the indices $d_j^*(N)$ yield simultaneous confidence intervals for DISP and SUPP that do not overlap, suggesting that SUPP is more important than DISP, as indicated by the point estimates. Simultaneous confidence intervals for the VI indices for DISP and INDEP do overlap, however, suggesting that the population importances of these two variables may not actually be different. The simultaneous confidence intervals are illustrated graphically in Figure 3.

Conclusions

This article has described a search for variable importance measures appropriate for logistic regression, motivated by earlier work on the relationship between corporate governance variables and the issuance of restricted shares. Two methods have been proposed, both of which are based on Pratt's (1987) axiomatically derived partition of R^2 for multiple linear regression, which can be generalized using the geometric interpretation described by Thomas et al. (1998). The first method uses a pseudo-R^2 measure for logistic regression proposed by McKelvey and Zavoina (1975) and Laitila (1993), which represents a logistic regression as the binary truncation of an unobservable dependent variable that is linearly related to the explanatory variables of interest.

This method yields a set of VI indices denoted $d_j^*(N)$ in the paper. The second method uses a representation of the maximum likelihood estimate of the logistic regression coefficients as a weighted least squares (WLS) regression, a representation exploited earlier by Pregibon (1981), Nordberg (1981) and Hosmer, Jovanovic and Lemshow (1989). A set of VI indices, denoted d_j^{WLS}, are then derived by applying a geometric analogue of Pratt's partitioning approach to the WLS version of R^2 based on this representation. Both sets of indices satisfy the property that they sum to one, which gives each index a meaningful scale, and they also share the property of additivity, namely that the importance of a subset of variables is equal to the sum of their individual importances, a property not shared by competing measures. A large synthetic dataset was constructed to mimic the actual data and was used to explore the small/medium sample properties of the two main methods. The indices $d_j^*(N)$ exhibited more stable small sample behaviour and were therefore used in the final analysis of variable importance.

In the analysis of the motivating case, the VI indices $d_j^*(N)$ were used to assign importances to three corporate governance factors that highlight difference in governance characteristics between firms with restricted share structure and other public firms without this structure. These variables were SUPP (suppression of shareholders interests), DISP (dispersion of ownership) and INDEP (board independence). A non-parametric bootstrap method was used on the RSF dataset to make statistical inferences on the importance measures.

Standard errors together with individual and simultaneous confidence intervals were estimated for each importance measure of the governance factors in the logistic regression model. The bias corrected and accelerated interval method (BCa) was employed to ensure good coverage performance of the confidence interval (Efron 1987; Platt, Henley and Yang 2000). The inferential analysis revealed that the most important contribution to the logistic regression, i.e., to the probability that a firm will issue restricted voting shares, is made by the variable SUPP. Although point estimates of importance suggest that variable DISP is more important than INDEP, examination of the simultaneous confidence intervals reveals that the importances of these two variables are not significantly different. It can be seen from the earlier results shown in Table 3 that the ranking suggested by the regression coefficients (which have identical scales because of the unit variances of the composite variables) and the Wald statistics are the same for the RSF variables as those suggested by the VI indices. This will not be the same in all situations, however, and occurs in this case because of the relatively small correlations between the explanatory corporate governance variables.

Though the development of the VI indices $d_j^*(N)$ described in this paper was motivated by an analysis of the RSF dataset, these indices and the general methodology can be applied to any logistic regression which can be modeled in terms of an underlying continuous response. Alternatively, if this assumption is deemed untenable in some situation, the alternative VI indices d_j^{WLS} based on the WLS representation can be used. It is important to note, however, that the examination of the properties of both sets of indices has been limited to a comparison with an empirically generated population. Further research involving simulation studies is needed to examine in detail the small and medium sample biases and confidence interval coverage rates of both sets of indices. In the meantime, however, the theoretical developments described in this paper provide a viable solution to the vexing problem of determining the relative importance of explanatory variables in a logistic regression analysis.

Authors' Note

A preliminary version of this paper was presented at the 2006 Administrative Science Association of Canada (ASAC) conference where it was awarded an honorable mention award in the Management Science track. The authors thank Professor Vijay Jog, Sprott School of Business, Carleton University, for permission to use the data that motivated the proposed methodology. The work of Thomas was supported by a grants from the Natural Sciences and Engineering Research Council of Canada, and from the Canadian research network MITACS (Mathematics of Information Technology and Complex Systems). Work of Zumbo was supported by grants from the Social Sciences and Humanities Research Council of Canada, and the Canadian Institutes for Health Research.

References

Azen, R., Budescu, D. V., and Reiser, B. (2001). Criticality of predictors in multiple regression. *British Journal of Mathematical and Statistical Psychology*, *54*, 201-225.

Azen, R., and Budescu D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*, 129-148.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*, 542-551.

Cameron, A. C. and Windmeijer, F. A. G. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of business and Economics Statistics*, *14*, 209-220.

Canty, A. and Ripley, B. (2006). Bootstrap R (S-Plus) functions. (http://cran.r-project.org/src/contrib/Descriptions/boot.html)

Cox, D. R. and Snell, E. J. (1989). *Analysis of binary data*. 2nd Edition, London: Chapman and Hall

Cragg, J. G. and Uhler, R. S. (1970). The demand for automobiles. *Canadian Journal of Economics*, *3*, 386-406.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. UK: Cambridge Press.

Drasgow, F. (1986). Polychoric and polyserial correlations. *The Encyclopedia of Statistics*, *7*, 68-74.

Efron, B. (1978). Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association*, *73*, 113-121.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association 82*, 171-185.

Efron, B. and Tibshrani, R. (1993). *An introduction to the bootstrap*. NY: Chapman and Hall.

Green, P.E., Carroll, J.D., and DeSarbo, W.S. (1978). A new measure of predictor variable importance in multiple regression. *Journal of Marketing Research*, *15*, 356-360.

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

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55-67.

Hosmer, D.W., Jovanovic, B. and Lemeshow, S. (1989). Best subsets logistic regression. *Biometrics*, 45, 1265-1270.

Hosmer, D. W., and Lemeshow, S. (2000). *Applied logistic regression*. Second Edition, NY: John Wiley & Sons, Inc.

Jog, V., Zhu, P. and Dutta, S. (2006). One share-one vote. Canadian Investment Review, Fall, 9-13.

Kruskall, W. (1987). Relative importance by averaging over orderings. *The American Statistician*, 41, 6-10.

Laitila, T. (1993). A pseudo-R^2 measure for limited and qualitative dependent variable models. *Journal of Econometrics*, *56*, 341-356.

McCullagh, P. and Nelder, J.A. (1989). *Generalized linear model*s, Second Edition. NY: Chapman and Hall.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zaremka (ed.), *Frontiers in Econometrics*. NY: Academic Press, p. 105-142.

McKelvey, R. D. and Zavoina (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*, 103-120.

Mittlbock, M. and Schemper, M. (1996). Explained variation for logistic regression. *Statistics in Medicine*, *15*, 1987-1997.

Nagelkerke, N. J. D. (1991). A note on general definition of the coefficient of determination. *Biometrika*, *78*, 691–692.

Nordberg, L. (1981) Stepwise selection of explanatory variables in the binary logit model. *Scandinavian Journal of Statistics*, *8*, 17-26.

Platt, R. W. Hanley, J. A., and Yang, H (2000). Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test. *Statistics in Medicine*, *19*, 313-322.

Pratt, J.W. (1987). Dividing the indivisible: using simple symmetry to partition variance explained. In T. Pukkila and S. Puntanen (Eds.), *Proceedings of the Second International Conference in Statistics* (p. 245-260). Tampere, Finland: University of Tampere.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, *9*, 705-724.

Thomas, D. R. (1992). Interpreting discriminant functions: a data analytic approach. *Multivariate Behavioral Research*, *27*, 335-362.

Thomas, D.R., Hughes, C.E. and Zumbo, B.D. (1998). On variable importance in linear regression. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, 45*, 253-275.

Thomas, D. R. and Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. *Journal of Educational and Behavioral Statistics*, *21*, 110-130.

Thomas, D. R. Zhu, P. and Decady Y. (2007). Point estimates and confidence intervals for variable importance in multiple linear regression. *Journal of Educational and Behavioral Statistics*, *32*, 61-91.

Windmeijer, F.A.G. (1995). Goodness-of-fit measures in binary choice models. *Econometric Review*, *14*, 101-116.