Regression-Based Decompositions: A New Tool for Managerial Decision-Making

Gary S. Fields
Cornell University
gary.fields@cornell.edu

November, 2003 *

Abstract

This paper shows how regression models can be supplemented by decomposition analyses to learn the relative importance of different explanatory factors. In regression analyses, the emphasis is on coefficients and statistical significance; in decompositions, it is on the information content of the variables in question.

I show how the explained portion of the regression (R-squared) can be decomposed into weights for each of the regressors. Moreover, if certain decomposition rules are accepted, then the weights given hold for a broad class of dispersion measures.

I then demonstrate how these methods of analysis have been applied to retention and performance of professionals at "Engineering Solutions" and profits at Borders. Action implications of the findings for each company are also presented.

For academics, regression-based decompositions hold great potential. For managers, it is very useful to know what makes a great deal of difference in their particular setting and, equally importantly, what does not.
`

* I am delighted to acknowledge the many helpful discussions I had with Jed DeVaro during the preparation of this paper. I am grateful to Ed Lazear and other colleagues at Stanford for useful comments on an earlier draft.

1. Introduction

The purpose of this paper is to show how regression-based decompositions, which earlier I devised for the analysis of income inequality, can be used to supplement the traditional information obtained from regressions in managerial decision-making. In the regression models that are at the core of empirical studies in much of our field, overwhelming emphasis is placed on the regression coefficients, their associated standard errors, and the consequent level of statistical significance. As users of regression analysis well know, the primary question asked by regression studies is, what is the effect on the dependent variable Y of a one unit (or percent) change in one independent variable X^j, holding other independent variables $X^1, X^2, \ldots X^{j-1}, X^{j+1}, \ldots X^K$ constant? The regression coefficients can be used to compare the effects of different variables $X^1, X^2, \ldots$, provided that these X's are measured in the same units.

A secondary question asked by regression studies is, how much of the variance in the dependent variable is accounted for by the X's taken together? Of course, the R^2 statistic provides the answer to this question.

Consider now a third question: How much of the variation in Y is accounted for by *each* of the independent variables? A possible answer to a question like this is, the information contained in X^1 is three times (or fives times or ten times) as important as the information contained in X^2, which in turn is twice as important (or as important or half as important) as the information contained in X^3. The answers are useful both to researchers who want to know what determines Y in different settings and to managers who want to know which X's they should manage and which they can safely ignore.

Before being able to come up with answers such as these, we need to ask, what metric provides a meaningful answer? There are several possibilities.

One is to compare the R^2's that come from simple regressions of Y on X^1, Y on X^2, etc. The advantage of this procedure is that it enables the explanatory power of individual independent variables to be compared with one another; the disadvantage is that other things have not been held equal.

This leads to a second way of assigning explanatory power to the various X's: running a standard one-way analysis of variance without interactions on $X^1, \ldots X^K$ and comparing the sums of squares associated with the different explanatory variables. The advantage of this method is that it enables the contribution of X^1 to be compared with that of X^2, etc. when all are included in the model. A disadvantage of analysis of variance is that the sum of squares explained by the model does not equal the sum of the sums of squares attributable to each explanatory variable.

Another concern arises with analysis of variance. By its very nature, the dispersion measure that is being analyzed is the variance. For many management applications– including the analyses of retention, performance, and profits presented in this paper - the variance is a natural dispersion measure to use. However, in other contexts – for example, in studying CEO pay dispersion – it might be interesting to use a different dispersion measure such as the 90:10 ratio or the Gini coefficient.

Here then is the challenge: Within the regression framework, is there a method for assigning explanatory power to the several independent variables that a) holds other things equal, b) decomposes in the sense that the contributions of the several independent variables sum to the contribution of the overall model, and c) allows for variation in the dependent variable to be gauged by an index other than variance? The answer, as I shall show, is "yes."

In the balance of this paper, I present a decomposition method that fulfills these requirements (Section 2). I then turn in Sections 3 and 4 to two applications:
i) understanding how an engineering firm was able to use these methods to gain insight into how to "keep its best," and ii) understanding how a major bookstore chain (Borders) was able to use these methods to learn which managerial competencies and activities contributed to profits, which made no appreciable difference, and which actually harmed profits. Section 5 concludes.

2. Decomposing the Dependent Variable into Contributions Made by Various Explanatory Factors

A. The Method

In a series of working papers ultimately published as Fields (2003), I devised a regression-based procedure for assigning weights to various independent variables in "accounting for" or "explaining" a dependent variable. The variation in the dependent variable is decomposed into as many components as there are factors.[1] This procedure constitutes a decomposition in the sense that the dispersion of the dependent variable, gauged for example by the variance, is broken down into a number of components such

[1] The method allows for two kinds of factors: "simple factors" that are represented by a single regressor and "composite factors" that are represented by two or more regressors. Examples of composite factors are 1) those that involve polynomials (e.g., potential experience, consisting of age-minus-schooling-minus-six and age-minus-schooling-minus- six squared) and 2) those that involve strings of dummy variables (e.g., race, consisting of dummy variables for black, white, and Asian).

that the whole is equal to the sum of its parts.[2] While the decomposition method developed was circulating as a working paper, the approach was used in income inequality studies around the world.[3]

The methods used in that literature are equally applicable to managerial studies. The essential results are the following.

Consider a standard regression equation of the form

$$Y = \beta^0 + \sum_{k=1}^{K} X^k \beta^k + \varepsilon. \tag{1}$$

When the equation is estimated in data, the resultant parameter estimates are

$$(\hat{\beta}^0 \hat{\beta}^1 ... \hat{\beta}^K) \tag{2}$$

and the calculated residuals are

$$\hat{\varepsilon}^i = Y^i - \hat{\beta}^0 - \sum_{k=1}^{K} X^{ik} \hat{\beta}^k, i = 1,...,n. \tag{3}$$

As described in the introduction, decomposition analysis is intended to supplement regression analysis by using the estimates in (2) and (3) to assign explanatory power to the several independent variables in a way that a) holds other things equal,

---

[2] The term "decomposition" has been used in this sense in many studies including the literature on inequality decomposition by factor components (e.g., Fei, Ranis, and Kuo, 1978; Pyatt, Chen, and Fei, 1980; and Shorrocks, 1982) and the literature decomposing differences in means between groups (Oaxaca, 1973; Blinder, 1974; Oaxaca and Ransom, 1994). However, "decomposition" has also been used in a more restrictive sense by Bourguignon (1979), who defines an inequality measure to be decomposable when the total inequality of a population can be broken down into a weighted average of (i) the inequality existing within subgroups of the population using that same inequality measure and (ii) the inequality existing between the subgroups. "Decomposition" is used here in the less restrictive sense.

[3] See Arcos (1996) for a study of Ecuador, Fields et al. (1998) for Bolivia, Sánchez and Núñez (1998) for Colombia, Fields and Mitchell (1999) for Taiwan, Ravallion and Chen (1999) for China, Fields and Yoo (2000) for South Korea, Contreras (2000) for Chile, Andersen (2001) for Bolivia, Redmond and Kattuman (2001) for the U.K. and Hungary , Heltberg (2001) for Vietnam, Yun (2002) for the U.S., Gang and Yun (2002) for East Germany, and Gindling and Trejos (2003) for Costa Rica.

b) decomposes in the sense that the contributions of the several independent variables sum to the contribution of the overall model, and c) allows for variation in the dependent variable to be gauged not only by the variance but by other indexes as well.

Criteria a) and b) can be fulfilled in the following manner. Given the regression equation (1) and estimates (2) and (3), let $s(X^k)$ denote the share of the variance of Y that is attributable to the k'th explanatory factor and let R^2 be the fraction of the variance that is explained by all of the X's taken together. Then, I show in Fields (2003) that the variance of Y can be decomposed as

$$\mathrm{var}(Y) = \sum_{k=1}^{K} \mathrm{cov}[X^k \hat{\beta}^k, Y] + \mathrm{cov}[\hat{\varepsilon}\ \ , Y] \ , \tag{4.a}$$

or, upon dividing through by var(Y),

$$100\% = \frac{\sum_{k=1}^{K} \mathrm{cov}[X^k \hat{\beta}^k, Y] + \mathrm{cov}[\hat{\varepsilon}\ \ , Y]}{\mathrm{var}(Y)} \equiv \sum_{k=1}^{K} s(X^k) \ + s(\hat{\varepsilon}), \tag{4.b}$$

where each "s-weight" $s(X^k)$ is given by

$$s(X^k) = \frac{\mathrm{cov}[X^k \hat{\beta}^k, Y]}{\mathrm{var}(Y)} \tag{4.c}$$

and the weight associated with the residual is given by

$$s(\hat{\varepsilon}) = \frac{\mathrm{cov}[\hat{\varepsilon}, Y]}{\mathrm{var}(Y)}. \tag{4.d}$$

It may be noted that when the last term in (4.b) is excluded, the remaining s-weights

$$\sum_{k=1}^{K} s(X^k) \ \equiv \frac{\sum_{k=1}^{K} \mathrm{cov}[X^{ik} \hat{\beta}^k, Y]}{\mathrm{var}(Y)}$$

sum exactly to R^2. Finally, expressing the $s(X^k)$'s in terms of their percentage contribution to R^2, we obtain the "p-weights"

$$p(X^k) \equiv \frac{s(X^k)}{R^2} \tag{4.e}$$

such that the $p(X^k)$'s sum to 100%. The results given in (4.a)-(4.e) provide a full and exact decomposition of the variance.[4]

The weights just derived take the variance as the measure of dispersion of Y. In some contexts, it may be desirable to allow the dispersion of Y to be measured by something other than the variance (i.e., criterion c) above), Fields (2003) applies a result developed by Shorrocks (1982) in another context to show that under six specified decomposition conditions (described in the appendix), the s-weights and p-weights given in (4.a)-(4.e) hold for *any* measure of dispersion that is continuous, symmetric and takes the value zero when all Y^i's are identical. These three properties are satisfied by virtually all dispersion measures including the coefficient of variation, income share of the richest x%, income share of the poorest y%, the M%/N% ratio (e.g., 90/10), Gini coefficient, Theil index, and Atkinson index.

This theorem provides an extremely powerful result. It says that as long as we agree on the conditions for carrying out the decomposition, we get *the same* s-weights and p-weights for a broad class of dispersion measures, including most of the commonly-used inequality measures.

---

[4] The decomposition here is "exact" in the sense that the variance of Y is decomposed exactly into the sum of components attributable to each regressor and the residual. By contrast, in standard analysis of variance, the sum of squares explained by the model is not the exact sum of the sums of squares attributable to each component.

B. Practical Implementation

Here is an important practical note. A Stata routine for implementing this decomposition was written by Stas Kolenikov of the University of North Carolina and is available online under the name gfields. It may be accessed by looking under STB and User-Written Programs in the Stata pull-down menu or by logging onto the internet at http://econpapers.hhs.se/software/bocbocode/s417004.htm. The gfields routine displays the s-weights given in (4.c) and (4.d). They can easily be transformed into proportions of R^2 ("p-weights") if you prefer.

C. Interpreting the Results

To illustrate the kinds of results yielded by this procedure, please refer to the first-stage multiple regressions in Tables 2, 5, and 7 and the corresponding decomposition results in Tables 3, 6, and 8. The results are discussed at length in Sections 3 and 4.

As a first example, consider regression (6) of Table 7 and the associated decomposition (6) in Table 8.[5] The variable being decomposed is district profits at Borders and the independent variables are shrinkage, consistency, etc. (The variables and results are detailed in Section 4.) From Table 7, we obtain regression coefficients and learn that 87% of the variation in profits is "explained" by the seven independent variables. Table 8 presents the corresponding decomposition results. The largest percentage contribution to profits is made by consistency, which accounts for 34.8% of the 87%, the next largest by optimistic, which accounts for 20.9%, and so on. In this way,

[5] Please note that the decompositions results are presented in the form of p-weights so that each factor's explanatory power is expressed as a percentage of R^2.

we learn that the least important of the variables with positive weight is results-driven, whose weight is 9.8%.

One variable – conflict management – exhibits a negative weight. Such a negative weight can be understood thus. The formula for the p-weight is

$$p(X^k) \equiv \frac{s(X^k)}{R^2} \tag{4.e}$$

where

$$s(X^k) = \frac{\mathrm{cov}[X^k \hat{\beta}^k, Y]}{\mathrm{var}(Y)}. \tag{4.c}$$

Recognize too that $\mathrm{cov}[X^k \hat{\beta}^k, Y] = \hat{\beta}^k \mathrm{cov}[X^k, Y]$ and that the regression of Y on X^k produces the simple regression coefficient $\beta_{X^k,Y} = \mathrm{cov}[X^k, Y] / \mathrm{var}(Y)$. Combining these four expressions, we have

$$p(X^k) = \frac{\hat{\beta}^k \beta_{X^k,Y}}{R^2} \; ; \tag{5}$$

that is, the numerator is the product of the multiple regression coefficient on X^k and the simple regression coefficient on X^k. A negative value for $p(X^k)$ arises whenever the two have opposite signs. In other words, an X^k contributes a positive weight in the multiple regression-based decomposition if and only if the *ceteris paribus* and *mutatis mutandis* effects of X^k go in the same direction.

It is also instructive to look at the pattern of p-weights reported in Table 6. Here, the dependent variable is retention of engineers, and fourteen possible explanatory variables are included. Six explanatory variables are found to have positive roles to play, but their contributions are quite different from one another. Eight other variables are found to have no role to play. These differing contributions would have been missed if only the regression coefficients in Table 5 had been examined.

Finally, it is also interesting to compare the results of the decomposition procedure with those that would have been gotten had partial R^2's been compared instead. For illustration, using the Borders data, suppose we work with just two explanatory variables, consistency and shrinkage. The partial correlation coefficient squared is a measure of the extent to which that part of the variation in the dependent variable which is not explained by the other predictors is explained by predictor *i*. These gauge the relationship between predictor *i* and the dependent variable, *y*, with the influence of the other variables in the regression equation eliminated:

$$r^2_{yi \cdot jl \ldots} = \frac{R^2_{y \cdot ijl \ldots} - R^2_{y \cdot jl \ldots}}{1 - R^2_{y \cdot jl \ldots}}$$

where

$R^2_{y \cdot ijl \ldots}$ = multiple R squared with predictor i

$R^2_{y \cdot jl \ldots}$ = multiple R squared without predictor i.

A regression of profit on consistency and shrinkage produced the following result:

$$\text{profit} = 8628606 - \underset{(480130)}{1608265}\ \text{consistency} + \underset{(216706)}{638853}\ \text{shrinkage}\ , \ R^2 = 0.34.$$

which in turn produces a partial R^2 of 0.19 for consistency and 0.23 for shrinkage. The main result from the partial R^2 method is that shrinkage is the more important variable than consistency. Note by the way that the partial R^2's do not sum to the total R^2.

How do these results compare to those from the decomposition methodology described above? The s-weights that we get from (4.c) are 0.191 for consistency and 0.145 for shrinkage. The corresponding p-weights from (5) are 56.8% for consistency and 43.2% for shrinkage. By this method, it is consistency, not shrinkage, that is the more important variable.

The conclusion, then, is that the preceding decomposition can make more than a quantitative difference; it can also make an important qualitative difference in deciding which explanatory variables are most important and which are less important determinants of the dependent variable.

D. <u>Summary</u>

In summary, then, here is what is to be done:

1. Using appropriate theory and past research, specify a regression equation of interest (eq. (1)).
2. Run the regression and derive "good" estimates of the regression coefficients and the residuals (eqs. (2) and (3)).
3. Run the decomposition using gfields in Stata. The program produces the s-weights (4.c) and (4.d).
4. If you prefer, calculate the p-weights by dividing the s-weights by R^2.

Let us turn now to two applications of this method in the management field.

3. Keeping the Best: Retention and Performance at "Engineering Solutions"

"After cracking the whip, employers who want to win the coming war for talent need to start giving their troops a compelling reason to stay."

Business Week, September 29, 2003, p. 92.

As this quote from Business Week indicates, the war for talent still rages.[6] In this section, I show how decomposition analysis was used to help a company learn about itself in its quest to win the war for talent.[7]

A. Data

At the start of 2002, a major U.S. human resources consultancy generously provided me with data on the engineers employed by one of their clients, here called "Engineering Solutions." To maintain confidentiality, nothing more was told to me about which industry the company operates in, how it generates its earnings, what specifically the engineers do, or how they are organized and managed.

Information was provided for 100 engineers hired by the company in 1996. Using the information on each of these engineers, I studied the drivers of retention, performance, and potential for promotion. Retention was gauged by whether the individual was still with Engineering Solutions at the start of 2002. Performance was rated on a five-point scale, ranging from "substantially exceeds expectations" down to "not meeting expectations." The scale was intended to be used so that all one-point increases represent the same improvement in performance. Typically the performance rating results from the manager's

---

[6] For earlier references, see Gubman, 1998; Johnson, 2000; Cappelli, 2000; Michaels, Handfield-Jones, and Axelrod, 2001; Tulgan, 2001; Welch, 2001; Collins, 2001; Watson-Wyatt, 2002; Ashby and Miles, 2002; and Bossidy and Charan, 2002.

[7] The results reported here are drawn from Fields (2002a, 2002b).

direct observation of the engineer's ability to carry out the technical requirements of a position. For space reasons, I omit the analysis of potential for promotion.

The available explanatory variables include objective data on several fixed characteristics including educational level, type of degree, age, gender, ethnicity, prior experience, and starting salary. I also have data on characteristics that changed over time: number of jobs held since joining the firm, number of separation days taken since joining the firm (a separation consists of a spell of four or more continuous days off other than vacation days, paid holidays, and sick days), and number of dependents. Last, I have subjective data on several personal attributes as rated by the most recent supervisor – communicates effectively, adapts to change, thinks creatively, and manages others effectively – each evaluated on a four-point scale ranging from "role model" to "does not display." The reason for collecting and using sensitive data on age, gender, and ethnicity is to assure that any decisions made by the company do not create or reinforce existing or pre-existing disparate impact.

B. Learning from Regression and Decomposition Analysis

Engineering Solutions has been very successful in retaining engineering talent. Fully 65% of the engineers hired in 1996 were still working with the company five years later. Despite this favorable record, the company hopes to do even better in retaining high-performing talent.

How did the engineers who stayed compare in quality with those who left? Table 1 shows the distributions of performance scores for the engineers hired in 1996 divided into two groups: the "stayers" who were still with the company after five years and the "leavers" who were not. We see that a much larger proportion of the stayers were performing at the top end of the scale. Meanwhile, most of the poor performers had left the organization. This evidence shows that the company has succeeded in retaining a particularly large share of the better performers while managing out the substandard ones.

What determines employee performance? Performance is measured on a five-point scale and each personal attribute on a four-point scale. Regression analysis shows that all of the personal attributes (creative thinking, adaptability to change, effective communication, and manages others) exhibit very powerful effects. The regression results (Table 2) reveal that an employee who is rated one point higher on each of the personal attributes performs a full point better – for example, moving from “meets expectations” to “exceeds expectations.” It would be very valuable for the company to be able to achieve such performance gains.

Turning now to decomposition analysis, when the variables known at the time of hiring and those learned afterwards are put together, the predominance of the personal attributes is confirmed (Table 3). These four attributes together account for 91% of the explained variation in performance, with thinking creatively being the most important. The only other variable that makes any appreciable difference to performance is number of jobs held in the company.

Turning now to the drivers of retention at Engineering Solutions, the statistical analysis presented in Table 4 reveals that some variables have large bivariate associations with retention, other variables are somewhat related to retention, and other variables make absolutely no difference at all. Specifically:

- The important (and statistically significant) drivers of retention are managing others effectively, adapting to change, communicating effectively, thinking creatively, and having fewer dependents.
- Other factors were found to have very small (and statistically insignificant) relationships with retention. These are ethnicity, gender, and prior experience.
- Finally, a number of other variables – type of engineer (chemical or mechanical), age, degree level, starting salary, and number of different jobs held in the firm since 1996 – were analyzed and found to make absolutely no difference to retention.

Which of the factors that matter, matter how much? In other words, if the company could learn only limited facts about their employees, which facts would be the best predictors of retention? The regression results for retention are reported in Table 5 and the corresponding decomposition results in Table 6. We see that when the engineers' characteristics are analyzed one at a time, managers' assessments of personal attributes clearly come out ahead. The top two personal attributes – manages others effectively and adapts to change – together provide 58% of the explanation for the turnover behavior of engineers, more than do all other factors taken together. Of the remaining variables, the next most important ones are number of separation days, followed by communicates effectively, number of dependents, and thinks creatively. Of the other variables, none accounts for even as much as 1% of retention.

I interpret these results as showing that Engineering Solutions exhibits considerable organizational learning. The variables that the company learned about over time enabled it to retain disproportionate numbers of a) the better performers,

b) employees with the personal attributes that make for greater success, c) employees who take fewer separation days, and d) employees with fewer dependent-related job issues. The company rarely dismissed professionals outright. I am told that those who were judged not to be performing to a high enough standard were encouraged to seek new situations, and most of them did.

C. Acting on the Findings

This information in the retention analysis is of enormous operational significance to Engineering Solutions. Based on these results, they can now conduct their hiring knowing which variables make an important difference to retention and performance in their own organization.

Turning to action implications, the statistical analysis led to four policy recommendations for the company:

First, personal attributes – communicating effectively, adapting to change, thinking creatively, and managing others - were shown to be the most important ones in explaining retention and performance at Engineering Solutions. Because these characteristics are the ones that matter the most, the company should obtain indicators of these attributes as early as possible – even at the time of initial screening - and select applicants accordingly.

Second, the number of separation days taken is a major negative factor in retention. Some high performers may be leaving to join companies that offer more flexible work-time arrangements. Engineering Solutions would do well to look seriously into what are usually called “family-friendly policies” but are better thought of as "person-friendly policies," in order to seek out cost-effective ways of increasing retention among loyal, high-performing employees who value a more flexible balance between work and other aspects of their lives than the company's current policy permits.

Third, the number of dependents is a major negative factor in retention. To the extent that the law permits, the company should look carefully at the engineers' family situations and see whether some problems might be mitigated, for example, by hiring a Dual Career Coordinator to help find suitable career opportunities for the spouses and partners of top engineers and other professionals.

Finally, like other employers, Engineering Solutions is forced by external pressures to pay higher salaries to engineers with graduate degrees than to those with bachelor's degrees only. This analysis shows, though, that engineers with graduate degrees are not significantly more likely to stay with the company than those with just bachelor's degrees and furthermore the performance scores of Ph.D.'s are not significantly higher than for master's and bachelor's degree holders. It would appear from this evidence that the hiring of Ph.D. degree holders at Engineering Solutions should be discontinued unless their employment can be justified on some other basis (such as the need to have a Ph.D. in order to be able to perform certain kinds of engineering tasks to a high professional standard).

4. Application: Profits at Borders

"The balanced scorecard . . . provides answers to four basic questions:

- *How do customers see us? [customer perspective]*
- *What must we excel at? [internal perspective]*
- *Can we continue to improve and create value? [innovation and learning perspective]*
- *How do we look to shareholders? [financial perspective]."*

Kaplan and Norton, 1992, p. 72.

Like most other economists, I learned the profit-maximization model and work with it regularly. I therefore believe strongly that scorecards should be *imbalanced* and that profits should receive central attention.

This section reports the highlights of a study of profitability at Borders. Further results are detailed in Fields and DeVaro (2003).

A. Data

In Summer 2003, Jed DeVaro (a Cornell colleague) and I obtained data from the Borders group on a number of outcome variables for the stores in each of their forty districts as well as various characteristics of the district managers and the stores in the district. From among the outcome variables, we chose to study profit in 2002. We then related profit to forty-one measures that included performance of the district on such variables as customer service and shrinkage (i.e., theft and damage) and evaluations of attributes of the district manager on such factors as being results-driven, persuasive, and optimistic.

Profit in millions of dollars was available from audited company records for each of the districts. Most of the performance measures were available for thirty-eight of the forty districts (one manager was not rated and one manager position was open), while many other measures were drawn from an assessment tool that had been administered only to the sixteen most recently-appointed district managers. The paucity of observations relative to variables necessitated a somewhat unusual approach to data analysis.

B. <u>Learning from Regression and Decomposition Analysis</u>

Specifically, DeVaro and I looked for variables that were significantly related in bivariate analysis to total profit in the district, profit per store, or profit as a percentage of sales. Then, taking profit as the dependent variable, we winnowed down the list of variables to those that exhibited statistical significance in the presence of other variables, looking at those variables that were available for thirty-eight districts. Three variables passed this test. Then, with these three as the base, we added in one at a time the four variables that had been found to be statistically significant among the sixteen district managers for whom they were available. The regression results are presented in Table 7.

The table includes seven explanatory variables that were found in bivariate analysis to be significantly related to profits:

- Shrinkage: Loss prevention from employee theft, customer theft, damage to the merchandise, etc.
- Consistency: An index composed of e-mail name collection café exception reports, customer service evaluations, customer service phone shop, the preferred program, special orders, corporate sales, and cost control.
- Informing: Keeping others well-informed by openly sharing information and maintaining a steady flow of communication.

- Conflict management: Addressing difficult situations directly and confronting conflict in a straightforward fashion.
- Results-driven: Achieving goals by challenging and pushing oneself to accomplish results at high levels.
- Feedback: Communicating directly and saying what is on one's mind in a candid, constructive, and direct way.
- Optimism: Presenting an upbeat image by being positive about future possibilities.

We see that six of these variables (shrinkage, informing, conflict management, results-driven, feedback, and optimism) are *positively* related to profits in the multivariate analysis. Of all of the specific skills that might have made a positive difference to Borders, these are the ones that are actually found to contribute to profits.

We see too that one of these variables (consistency) is *consistently and negatively* related to profits. This is quite remarkable: districts in which managers do better on consistency are *less* profitable. Moreover, the consistency variable is consistently the one with *the highest weight in the decomposition analysis* (Table 8). DeVaro and I interpret these findings thus:

> The finding that higher ratings on consistency are associated with lower district profit might appear surprising, since it seems hard to fathom that high performance on any of the dimensions of consistency could actually *hurt* profit. But given that inputs of managerial time are scarce, any effort devoted to improving consistency is invested at the expense of something else. It is plausible, and indeed the results of this study suggest it is likely, that other activities such as loss prevention should receive higher priority. To the extent that managers waste time improving "consistency" at the expense of more important managerial activities, profit may suffer.

C. Acting on the Findings

Upon being presented with these findings in September 2003, Borders took three steps. First, they became aware that the current definition of "consistency" has limitations, which they plan to address in the coming months. Second, they found that many of the manager attributes that made a positive difference to profits had been obtained from an assessment tool administered by a psychological testing service. Accordingly, they decided that the costs of using this tool to help select managers were well worth it, and so they are continuing with this practice. And third, they reconfirmed the potential benefits of training programs for district managers and store managers around the six items (shrinkage, informing, conflict management, results-driven, feedback, and optimism) that made a significant positive difference to profits. The first of these programs is expected to be launched shortly.

5. Conclusion

This paper has shown how regression models can be supplemented by decomposition analyses to learn the relative importance of different explanatory factors. In regression models, the emphasis is on coefficients and statistical significance; in decompositions, it is on the information content of the variables in question.

In Section 2, I showed how the explained portion of the regression (R-squared) can be decomposed into weights for each of the regressors. Moreover, if certain decomposition rules are accepted, then the weights given hold for a broad class of dispersion measures.

Sections 3 and 4 then presented applications of these methods to analysis of retention and performance at "Engineering Solutions" and profits at Borders. Action implications of the findings for each company were also suggested.

For academics, regression-based decompositions hold great potential. For managers, it is very useful to know what makes a great deal of difference in their particular setting and, equally importantly, what does not.

Appendix

The Six Conditions on the Decomposition

In the text, I stated that if six decomposition conditions are accepted, then the contribution of each X^k to the dispersion (or inequality) of Y can be expressed in terms of (4.a)-(4.e) for *any* inequality measure that is continuous, symmetric, and for which

$I(\mu \;\; \mu \;\; . \; . \; . \; \mu) = 0$. Denoting $X^k \hat{\beta}^k$ by Y^k, the six conditions are:

Condition 1: (Number of Components) The inequality measure I(Y) is to be divided into K components, one for each regressor, denoted $S_k(Y^1, \ldots, Y^K; K)$.

Condition 2: (a) (Continuity) Each S_k is continuous in Y^k. (b) (Symmetric Treatment of Factors) If $\pi_1, \ldots, \pi_k$ is any permutation of 1, . . , K, then $S_k(Y^1, \ldots, Y^K; K) = S_{\pi k}(Y^{\pi 1}, \ldots, Y^{\pi K}; K)$.

Condition 3: (Independence of the Level of Disaggregation) The amount of inequality accounted for by any one factor S_k does not depend on how the other factors are grouped: $S_k(Y^1, \ldots, Y^K; K) = S_k(Y^k, Y)$.

Condition 4: (Consistent Decomposition) The contributions S_k sum to the overall amount of inequality, viz., $\sum_k S_k(Y^1, \ldots, Y^K; K) = I(Y)$.

Condition 5: (a) (Population Symmetry) If P is any n x n permutation matrix, $S(Y^k P, Y P) = S(Y^k, Y)$; (b) (Normalization for Equal Factor Distribution) If all income recipients have the same value for the k'th factor, then the share of inequality accounted for by that factor $S(\mu_k e, Y) = 0$ for all μ_k.

Condition 6: (Two Factor Symmetry) Suppose the distribution of Y^2 is simply a permutation of that for Y^1. Then if those were the only two components of Y^k, Y^1 and Y^2 should receive the same value in the decomposition. Thus, for all permutation matrices P,

$S(Y^1, Y^1 + Y^1P) = S(Y^1P, Y^1 + Y^1P)$.

In my view, all of these conditions are quite acceptable. I therefore conclude that the decomposition procedure given in (4.a)-(4.e) is a powerful way of determining weights to assign to the various regressors in a linear model.

References

Andersen, L. (2001). *Low Social Mobility in Bolivia: Causes and Consequences for Development.* Unpublished Working Paper, Kiel Institute of World Economics, Germany.

Arcos, X. R. (1996). *Descomposición de la Desigualdad del Consumo en Ecuador*. Masters Thesis, Cornell University, Ithaca, New York.

Ashby, M.D. & Miles, S.A. (2002). *Leaders Talk Leadership.* (New York: Oxford University Press).

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8, 436-455.

Bossidy, L. & Charan, R. (2002). *Execution.* (New York: Crown Business).

Bourguignon, F. (1979). Decomposable Income Inequality Measures. *Econometrica.* 47:4, 901-920.

Cappelli, P. (2000). *The New Deal at Work.* (Boston: Harvard Business School Press).

Collins, J. (2001). *Good to Great.* (New York: Harper Collins).

Contreras, D. (2000). *Explaining Wage Inequality in Chile: Does Education Really Matter?* Unpublished Working Paper, Department of Economics, Universidad de Chile, Chile.

Fei, J. C. H., Ranis, G., & Kuo, S. W. Y. (1978). Growth and the Family Distribution of Income by Factor Components. *Quarterly Journal of Economics*, 92(1), 17-53.

Fields, G. S. (2002a). *Keeping Our Best: Econometric Analysis of Retention and Performance at Engineering Solutions.* Unpublished Working Paper, Cornell University.

Fields, G. S. (2002b). *Predicting Potential for Promotion: How the Data in Human Resource Information Systems Can be Used to Help Organizations Gain Competitive Advantage.* Unpublished Working Paper, Cornell University.

Fields, G. S. (2003). Accounting for Income Inequality and Its Change: A New Method, with Application to the Distribution of Earnings in the United States, *Research in Labor Economics,* Vol. 18.

Fields, G. S. & DeVaro, J. L. (2003). *Doing the Job Right Versus Doing the Right Job: What Helps and Hurts Profits at Borders?* Unpublished Working Paper, Cornell University, 2003.

Fields, G. S., Leary, J., López Calva, L., & Pérez de Rada, E. (1998). *Education's Crucial Role in Explaining Labor Income Inequality in Urban Bolivia*. Harvard Center for International Development, Available online at http://www.cid.harvard.edu/hiid/658.pdf.

Fields, G. S., & Mitchell, J. (1999). Changing Income Inequality in Taiwan: A Decomposition Analysis. In T. N. Srinivasan and G. Saxonhouse (Eds.), *Development, Duality, and the International Regime: Essays in Honor of Gustav Ranis*. Ann Arbor: University of Michigan Press.

Fields, G. S., & Yoo, G. (2000). Falling Labor Income Inequality in Korea's Economic Growth: Patterns and Underlying Causes. *Review of Income and Wealth*, 46(2), 139-159.

Gindling, T. H., & Trejos, J. D. (2003). *Accounting for Changing Earnings Inequality in Costa Rica in the Final Quarter of the 20th Century*. Unpublished Working Paper, University of Maryland, Baltimore County, and Universidad de Costa Rica.

Gubman, E. (1998). *The Talent Solution.* New York: McGraw Hill.

Heltberg, R. (2001). *Analyzing Inequality Using Income Regressions: Vietnam, 1992-97*. Unpublished Working Paper, University of Copenhagen, Denmark.

Johnson, M. (2000). *Winning the People Wars*. (London: Prentice Hall).

Kaplan, R. S. & Norton, D. P. (1992). The Balanced Scorecard: Measures that Drive Performance, *Harvard Business Review,* January-February.

Michaels, E., H. Handfield-Jones, & B. Axelrod (2001). *The War for Talent.* (Boston: Harvard Business School Press).

Oaxaca, R. L. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.

Oaxaca, R. L., & Ransom, M. R. (1994). On Discrimination and the Decomposition of Wage Differentials. *Journal of Econometrics*, 61(1), 5-21.

Pyatt, G., Chen, C., & Fei, J. (1980). The Distribution of Income by Factor Components. *Quarterly Journal of Economics*, 95(3), 451-473.

Ravallion, M., & Chen, S. (1999). When Economic Reform is Faster than Statistical Reform: Measuring and Explaining Income Inequality in Rural China. *Oxford Bulletin of Economics and Statistics*, 61(1), 33-56.

Redmund, G., & Kattuman, P. (2001). *Employment Polarisation and Inequality in the UK and Hungary*. Unpublished Working Paper, University of New South Wales and Cambridge University, U.K.

Sánchez Torrez, F., & Méndez, J. N. (1998). Descomposición de la Desigualdad del Ingreso Laboral Urbano: 1976-1997. In F. S. Torrez (Ed.), *La Distribución del Ingreso en Colombia*. Unpublished Working Paper, Departamento Nacional de Planeación, Bogotá.

Shorrocks, A. F. (1982). Inequality Decomposition by Factor Components. *Econometrica*, 50(1), 193-211.

Shorrocks, A. F. (1983). The Impact of Income Components on the Distribution of Family Incomes. *Quarterly Journal of Economics*, 98(2), 311-326.

Tulgan, B. (2001). *Winning the Talent Wars*. (New York: Norton).

Watson-Wyatt (2002). *Strategic Rewards Survey.* (Washington, DC: Watson-Wyatt).

Welch, J. (2001). *Jack: Straight from the Gut.* (New York: Warner Business Books).

Yun, M. (2002). *Earnings Inequality in USA, 1961-1999: Comparing Inequality Using Earnings Equations*. Unpublished Working Paper, University of Western Ontario, processed.

Table 1.
Engineering Solutions:
Performance of Stayers Compared with Leavers.

Variable	Engineers Still with the Company as of January, 2002 ("Stayers") n =65	Engineers No Longer with the Company ("Leavers") n =35	Percentage of Those With That Characteristic Still With the Company as of January, 2002
Substantially exceeds expectations	8%	0%	100%
Exceeds expectations	43%	9%	90%
Meets expectations	43%	23%	78%
Meets only some expectations	3%	51%	10%
Not meeting expectations	3%	17%	25%

Table 2.
Engineering Solutions: Regressions Explaining Performance Among the Stayers.
(Standard Errors in Parentheses)
(n=65)

Independent Variable	**Regression Coefficients**
Ph.D.	0.029
	(0.225)
M.S.	-0.026
	(0.222)
Chemical Engineer	0.329*
	(0.192)
Age	-0.000
	(0.012)
Asian	0.343*
	(0.182)
Black	0.076
	(0.229)
Male	-0.153
	(0.179)
Prior Experience	0.225
	(0.208)
Starting Salary	0.0000006
	(0.000009)
Number of Dependents	0.034
	(0.081)
Separation Days	0.002
	(0.005)
Effective Communication	0.216*
	(0.127)
Adapts to Change	0.259**
	(0.106)
Creative Thinking	0.478***
	(0.114)
Manages Others Effectively	0.183
	(0.118)
Number of Jobs Held	0.104
	(0.076)
Constant	-3.712***
	(0.727)
R-squared	0.67

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.
Engineering Solutions: Decomposition Analysis of the Factors Contributing to the Performance of Engineers.

Independent Variable	Percentage Contribution of the Variable in a Multivariate Decomposition (p-weights)
Thinks creatively	42.1%
Adapts to change	18.4%
Communicates effectively	15.7%
Manages others well	14.7%
Number of jobs held in the company	9.8%
Ethnicity	3.8%
Prior experience	1.8%
Number of separation days	1.2%
Number of dependents	1.2%
Starting salary	0.0%
Age	0.0%
M.S.	-0.3%
Ph.D.	-0.5%
Gender	-1.3%
Chemical/mechanical engineer	-6.7%
Total	100%

Table 4.
Engineering Solutions:
Bivariate Analysis of Factors Contributing to the Retention of Engineers.
(n=100)

Independent Variable	**Percentage Contribution of the Variable in a Bivariate Logistic Regression**
Manages others effectively (+)	44.7% **
Adapts to change (+)	39.2% **
Communicates effectively (+)	37.4% **
Thinks creatively (+)	26.8% **
Number of separation days taken (-)	18.4% **
Number of dependents (-)	5.3% **
Ethnicity (+ for Asian/Pacific)	1.7%
Male/female (+ for male)	0.4%
Prior experience (+)	0.4%
Mechanical/chemical engineer	0.1%
Age	0.1%
Degree level	0.0%
Starting salary	0.0%
Number of jobs held in the company	0.0%

Notes to Table 4:

Variables marked by a + raise retention. Variables marked by a – lower retention.

The percentage contributions are the pseudo-R^2's obtained from bivariate logits of stay/leave on the independent variable in question. Variables statistically significant at the .01 level are marked by **. No other variables are found to be statistically significant in this analysis.

Table 5.
Engineering Solutions: Regressions Explaining Retention.
(Standard Errors in Parentheses)
(n=100)

Independent Variable	**Regression Coefficient**
Effective Communication	0.063
	(0.052)
Adapts to Change	0.101**
	(0.045)
Creative Thinking	0.043
	(0.045)
Manages Others Effectively	0.175***
	(0.046)
Separation Days	-0.005***
	(0.001)
Number of Dependents	-0.088***
	(0.028)
Asian	-0.054
	(0.077)
Black	-0.041
	(0.086)
Male	-0.011
	(0.075)
Prior Experience	0.076
	(0.074)
Chemical Engineer	0.088
	(0.074)
Masters Degree in 1996	0.086
	(0.073)
Ph.D. in 1996	0.061
	(0.078)
Starting Salary	0.0000002
	(0.00000004)
Number of Jobs Held	-0.033
	(0.024)
Age	-0.001
	(0.005)
Constant	-0.160
	(0.282)
R-squared	0.71

* significant at 10%; ** significant at 5%; *** significant at 1%

**Table 6.
Engineering Solutions: Decomposition Analysis of the
Factors Contributing to the Retention of Engineers.
(n=100)**

Independent Variable	Percentage Contribution of the Variable in a Multivariate Decomposition (p-weights)
Manages others effectively	38.9%
Adapts to change	18.9%
Number of separation days	17.5%
Communicates effectively	10.9%
Number of dependents	7.6%
Thinks creatively	6.4%
Prior experience	0.7%
Chemical engineer	0.5%
Starting salary	0.0%
Male	-0.1%
Age	-0.1%
Degree level in 1996	-0.2%
Number of Jobs Held	-0.4%
Ethnicity	-0.7%
Total	100%

Table 7.
Borders: Regressions of District Profits on Managers' Attributes.
(Standard errors in parentheses)

Independent Variable	**Regression Coefficients**					
	(1)	(2)	(3)	(4)	(5)	(6)
Shrinkage	583,107**	389,577	267,578	553,552	279,421	377,812
	(217,507)	(383,987)	(486,189)	(549,417)	(323,845)	(435,959)
Consistency	-1,794,971***	-2,208,419***	-2,383,767***	-1,834,785***	-1,881,074***	-2,130,403***
	(475,420.744)	(365,867)	(517,245)	(495,850)	(370,044)	(390,170)
Informing	1,432,103**	1,959,987**	1,572,310	1,974,712*	2,179,754**	1,602,372
	(593,049)	(845,403)	(895,116)	(933,850)	(759,877)	(904,167)
Conflict management		58,082***				-14,304
		(11,575)				(27,854)
Results-driven			62,427**			38,090
			(20,916)			(22,899)
Optimistic				51,623*		45,234
				(23,309)		(27,787)
Feedback					46,169***	40,835**
					(12,256)	(11,703)
Constant	4,347,649*	444,411	2,990,742	-1,414,741	77,979	-1,730,640
	(2,412,436)	(2,689,071)	(3,442,819)	(4,452,986)	(3,240,021)	(3,164,373)
R-squared	0.43	0.74	0.73	0.68	0.75	0.87
n	38	15	15	15	15	15

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8.
Borders: Decomposition Analysis of the Factors Contributing to District Profits.

Independent Variable	**Percentage Contribution of the Variable in a Multivariate Decomposition (p-weights)**					
	(1)	(2)	(3)	(4)	(5)	(6)
Shrinkage	30.6%	15.3%	10.9%	10.7%	22.0%	11.8%
Consistency	49.2%	45.2%	50.7%	37.6%	38.2%	34.8%
Informing	20.6%	21.0%	17.5%	22.8%	21.5%	13.7%
Conflict management		18.5%				-3.7%
Results-driven			20.8%			9.8%
Optimistic				28.9%		20.9%
Feedback					18.3%	12.6%
Total	100%	100%	100%	100%	100%	100%
n	38	15	15	15	15	15