What Does $R^2$ Mean in ANOVA?

Richard P. Rumelt
The Anderson School at UCLA

April 10, 1998

Schmalensee [1985] was the first to apply variance components methods to measure the “importance” of industry in accounting for the variance in profitability among business units. Since then, studies by Rumelt [1991], Roquebert et. al. [1996], and McGahan and Porter [1997] have treated this issue in more detail and with a variety of methods. The purpose of this note is to explain the basic reason for choosing variance components decomposition over ANOVA. Put differently, this note explains the problems that arise when one attempts to use incremental contributions to $R^2$ as a proxy for “importance” or variance components.

1 Nested Models

Consider the following nested model:

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk},$$  \hspace{1cm} (1)

where $i = 1, 2, \ldots a$, $j = 1, 2, \ldots b_i$, and $k = 1, 2, \ldots c_{ij}$. The total number of observations is

$$N = \sum_{i=1}^{a} \sum_{j=1}^{b_i} c_{ij}. \hspace{1cm} (2)$$

This model posits two levels of classification. The lower B-level ($\beta$ effects) is nested within the higher A-level ($\alpha$ effects). There are $a$ groups in the A-level of classification and within group $i$ there are $b_i$ sub-groups in the lower B-level of classification. Finally, within subgroup $j$ of group $i$ there are $c_{ij}$ observations of the dependent variable $y_{ijk}$. In the context of business-unit performance, the B-level would be the business-unit and the A-level the industry (or the corporate parent). In another context, the B-level might be students and the A-level might be classes, each observation being a test score.
1.1 Notation
Regression estimation of classification models is normally accomplished by using dummy (0,1) variables. Writing (1) as a dummy-variable regression equation would require a more complex and less intuitive notation. The purpose of the dummy variables is to “select” the coefficients that are active in any particular observation. In (1) the notation shows only the selected coefficients for the observation in question.

A second useful notation is to represent a sum over a subscript with a dot. Thus, the expression $\sum_{i=1}^{a} b_i$ can be replaced by $b_\cdot$ and (2) can be more compactly expressed as $N = c_\cdot\cdot$.

1.2 Estimability
An important theorem about linear models concerns what is estimable via regression (or ANOVA) methods. Regression models construct estimates of model parameters as linear combinations of observations. Theorem 1 allows one to determine whether or not a particular function is estimable using regression.

**Theorem 1** The expected value of any observation is estimable and so are all linear combinations thereof.


That is, $\mu + \alpha_i + \beta_{ij}$ is estimable for all valid $i, j$. From the fact that all linear combinations of expected values of observations are estimable, we know that

$$E_\chi = \sum_{i=1}^{a} \sum_{j=1}^{b_i} \lambda_{ij}(\mu + \alpha_i + \beta_{ij})$$

is estimable for any $\lambda_{ij}$. Rewriting this as

$$E_\chi = \mu \lambda_{\cdot\cdot} + \sum_{i=1}^{a} \alpha_i \lambda_{i\cdot} + \sum_{i=1}^{a} \sum_{j=1}^{b_i} \beta_{ij} \lambda_{ij}$$

permits intuitive proofs of the following observations:

**Observation 1** $\mu$ is not estimable.

*Proof:* For $\mu$ to be estimable the $\lambda_{ij}$ must be chosen such that $E_\chi = \mu$. This requires $\lambda_{\cdot\cdot} = 1$ and, at a minimum, $\sum_{ij} \beta_{ij} \lambda_{ij} = 0$ for all $\beta_{ij}$. These two conditions are incompatible, so $\mu$ is not estimable. *Q.E.D.*

**Observation 2** $\mu + \alpha_p$ is not estimable.
Proof: To select $\mu + \alpha_p$ we require $\lambda_{p\ast} = 1$ and $\lambda_{i\ast} = 0$ for $i \neq p$. This reduces (4) to

$$E_\lambda = \mu + \alpha_p + \sum_{j=1}^{b_p} \lambda_{pj}\beta_{pj}.$$ 

Since the $\lambda_{pj}$ must sum to 1, the last term cannot be eliminated for arbitrary $\beta_{pj}$. Hence, $\mu + \alpha_p$ is not estimable. \(Q.E.D.\)

**Observation 3** $\alpha_p$ is not estimable.

*Proof:* The proof parallels that of Observation 2. \(Q.E.D.\)

These observations establish that one cannot estimate the overall mean or the top-level class effects in a two-level nested model. The next observation establishes the even stronger statement that even the “contrasts” between A-level effects are not estimable.

**Observation 4** $\alpha_p - \alpha_q$ $(p \neq q)$ is not estimable.

*Proof:* Assume that $\lambda_{ij}$ are chosen such that $\lambda_{p\ast} = 1$, $\lambda_{q\ast} = -1$, and $\lambda_{i\ast} = 0$ for all $i \ni \{p, q\}$. Then

$$E_\lambda = \alpha_p - \alpha_q + \sum_{j=1}^{b_p} \lambda_{pj}\beta_{pj} - \sum_{j=1}^{b_q} \lambda_{qj}\beta_{qj}.$$ 

The terms in $\beta_{pj}$ and $\beta_{qj}$ cannot, in general, be eliminated. Hence A-level contrasts are not estimable. \(Q.E.D.\)

The B-level contrasts are estimable. **In a nested model, only the contrasts among the lowest level classes are estimable.**

**Observation 5** $\beta_{sp} - \beta_{sq}$ is estimable for $p \neq q$.

*Proof:* Set $\lambda_{sp} = 1$ and set $\lambda_{sq} = -1$. Set all other values of $\lambda_{ij} = 0$. Evaluation of (4) yields $E = \beta_{sp} - \beta_{sq}$. \(Q.E.D.\)

2 Regression

In order to simplify and make the exposition more concrete I shall work with the example data set given in Table 1. In this data $a = 2, b_1 = 3, b_2 = 2, c_{11} = c_{12} = c_{13} = c_{22} = 2$, \(\ldots\)}
and $c_{21} = 3$. Also, $N = 11$. It is useful to examine how this problem would be analyzed using dummy variables in an OLS regression. The full $X$ matrix would appear as shown in Figure 1. The first dummy variable represents the mean and is always 1. Restrict attention to just the A-level effects: notice that the dummy variables for the two A-classes sum to the first column. That is, the first three columns are linearly dependent—these effects cannot be independently estimated. The regression cannot be performed because the matrix $X'X$ is singular. The standard approach to this problem is to set one of the A-class effects to zero—say A2. The resultant matrix (omitting the B-level effects) is shown in Figure 2.

Performing this regression yields $M^o = 8.6$ and $A1^o = 20.23$. It is important to understand the relationship between these dummy variables and the true model. Regression on dummy variables in 1-way or simple nested models is equivalent to the computation of group averages. Consequently, the estimate $M^o$ is just the average of the last 5 obser-
vatons and the estimate $M^o + A1^o$ is the average of the first 6 observations:

\[
M^o = \alpha^o + \frac{[3\beta_{21}^o + 2\beta_{22}^o]}{5} \\
M^o + A1^o = \alpha^o + \frac{[2\beta_{11}^o + 2\beta_{12}^o + 2\beta_{13}^o]}{6}
\]

so that

\[
A1^o = \alpha^o - \alpha^o + \frac{[2\beta_{11}^o + 2\beta_{12}^o + 2\beta_{13}^o]}{6} - \frac{[3\beta_{21}^o + 2\beta_{22}^o]}{5}
\]  

(5)

Fitting the A-level model generates estimates that unavoidably mix A-level and B-level effects.

2.1 The ANOVA Table

The following values are required for the ANOVA analysis:

\[
T_0 = \sum_{ijk} y_{ijk}^2 = 8628.0 \\
T_m = \bar{y}_{***}/N = 4241.5 \\
T_A = (\sum_i y_{i**})^2/c_i = 5358.0 \\
T_{AB} = \sum_{ij} y_{ij*}^2/c_{ij} = 8200.5
\]

These sums of squares are the center of the ANOVA analysis. ANOVA is not a special form of analysis—it is the same table of sum-squares that would be produced by OLS
dummy-variable regression. However, when class-effects models are used, especially when there are many levels, the estimated class-effects are often not reported. Instead, the impact of the whole classification level is reported on the ANOVA table.

$T_0$ is the total sum of squares: each observation is squared and then all are summed. $T_m$ is the “sum of squares due to the mean,” which is what the total sum of squares would be if each observation were equal to the average observation $y_{\bullet\bullet}/N$. Logically, the regression problem is to explain the difference between $T_0$ and $T_m$ with predictor variables. $T_A$ is the “sum of squares due to the A-level classification,” which is what the total sum-of-squares would be if each observation were equal to its predicted value using the fitted A-level effects. In a nested model, these predicted values are simply the class averages. Similarly, $T_{AB}$ is the “sum of squares due to the B-level classification,” which is what the total sum-of-squares would be if each observation were equal to its B-level class average. The notation $T_{AB}$ includes the $A$ term to remind us that the B-classes are nested within the A-classes.

Table 2 displays the analysis of variance table for fitting just the A-level effects. It is important to understand what the lines on this table mean.

Table 2: Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>sum-squares</th>
<th>df</th>
<th>Mean-Square</th>
<th>F</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R(\mu)$</td>
<td>4241.46</td>
<td>1</td>
<td>4241.46</td>
<td>11.67</td>
<td></td>
</tr>
<tr>
<td>$R(\alpha</td>
<td>\mu)$</td>
<td>1116.51</td>
<td>1</td>
<td>1116.51</td>
<td>3.07</td>
</tr>
<tr>
<td>$SSE$</td>
<td>3270.03</td>
<td>9</td>
<td>363.34</td>
<td></td>
<td>0.745</td>
</tr>
<tr>
<td>$SST_m$</td>
<td>4386.55</td>
<td>10</td>
<td>438.66</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$SST$</td>
<td>8628.00</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the other lines on the table Searle’s [1971] notation has been used. The first line on the table, $R(\mu)$, shows the reduction in total sum-squares due to fitting $\mu$. Numerically, $R(\mu)$ is what $SST$ would have been had each observation been the average. That is, $R(\mu) = T_m = N(y_{\bullet\bullet}/N)^2 = y_{\bullet\bullet}^2/N$. Next, fitting the model $\mu + \alpha_i$ produces a “reduction” in sum-squares of $R(\mu, \alpha) = T_A$. What is reported on the ANOVA table is the incremental reduction in sum-squares due to fitting the A-effects after the mean. This is $R(\alpha|\mu) = R(\mu, \alpha) - R(\mu) = T_A - T_m$. Finally, $SSE$ is the “error” sum-squares and is simply the residual: $SSE = SST - R(\mu, \alpha) = T_0 - T_A$. 

6
In many texts and statistical reports, \( R(\mu) \) will be represented by the symbol \( SSM \) and the total sum-squares due to fitting the regression will be \( SSR \) and, adjusted for the mean, \( SSR_m \). The \( R() \) notation is clearer when several class variables are being used and one wants to keep careful track of the differences between the total regression and incremental steps in fitting the model.

Note that the coefficient of determination \( R^2 \) is \( 1 - SSE/SST_m \). That is, it is the proportion of the sum-squares, adjusted for the mean, which is explained by the regression.

The degrees-of-freedom column shows the number of parameters estimated for regression elements \( (R(\mu) \) and \( R(\alpha|\mu)) \). Note that \( df(\alpha|\mu) = df(\alpha, \mu) - df(\mu) \), and that the number of parameters estimated is limited by the rank of the corresponding \( X'X \) matrix, and is not necessarily the number of levels in the classification. In the other lines of the table the degrees-of-freedom column shows the number of observations less the number of parameters estimated to arrive at the sum-squares. The Mean-Square column is the sum-squares divided by the corresponding degrees-of-freedom. For the \( SST_m \) line, \( MST_m = SST_m/(N - 1) \) shows the sample variance. For \( SSE \) the Mean-Square column shows the estimate of the error variance \( \hat{\sigma}^2 = MSE = SSE/(N - a) \).

The F-statistics are the ratios of the Mean-Squares for that line to \( MSE \). Under the null hypothesis that all the effects added by a line on the ANOVA table are zero, the F-statistic has a central \( F \) distribution with parameters given by the degrees-of-freedom for that line and the degrees-of-freedom for the \( SSE \) line. Thus, the F-statistic shown for \( R(\alpha|\mu) \) has parameters \((1,9)\).

### 2.2 Adding the B-level Effects

Referring to Figure 1, we can see several linear dependancies when the B-level effects are also considered. All of the B-effects sum to the first column, \( B1 + B2 + B3 \) sum to the column for A1 and \( B4 + B5 \) sum to the column for A2. Thus, \( A1 + B4 + B5 \) sum to the first column. The standard approach in this case is to eliminate all A-level dummy variables and one of the B-level dummies, say B5. The resultant \( X \) matrix is shown in Figure 3. (One could just as easily omit columns \( \mu \), A1, and A2, leaving only columns corresponding to the B-level classes).

Performing this regression yields \( M^o = 2 \), \( B1^o = 49.5 \), \( B2^o = -1.5 \), \( B3^o = 32.5 \), and \( B4^o = 11 \). At this point, it is useful to note what these dummy variables estimate. Again, we can deduce the parameter estimates from the fact that \( M^o \) is the average of the last 2 observations, \( M^o + B1^o \) is the average of the first 2 observations, etc. Thus,

\[
M^o = \mu^o + \alpha_2^o + \beta_{22}^o \\
B1^o = \alpha_1^o - \alpha_2^o + \beta_{11}^o - \beta_{22}^o \\
B2^o = \alpha_1^o - \alpha_2^o + \beta_{12}^o - \beta_{22}^o
\]
Figure 3: B-level Matrix of Independent Variables

<table>
<thead>
<tr>
<th>M</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ B3^o = \alpha_1^o - \alpha_2^o + \beta_{13}^o - \beta_{22}^o \]
\[ B4^o = \beta_{21}^o - \beta_{22}^o. \]

Except for B4, the dummy variable estimates combine estimates of the A-level and B-level effects. By taking differences we can obtain pure contrasts between the \( \beta \)'s within each A-level class:

\[ B1^o - B2^o = \beta_{11}^o - \beta_{12}^o \]
\[ B1^o - B3^o = \beta_{11}^o - \beta_{13}^o. \]

This result corresponds to Observation 5. Corresponding to Observation 4, note that no manipulations of the above expressions can yield pure estimates of the \( \alpha \) contrasts.

The ANOVA table is given in Table 3. This table adds to Table 2 the line \( R(\alpha, \beta|\mu) \). The sum-squares for this line is \( T_{AB} - T_m \). It tests the full model—which has been shown to be simply the B-level effects. The sum-squares from modeling the B-level effects are 3959.05 and the resultant \( R^2 = 0.903 \). We can add an additional level of analysis, however, by examining the incremental addition to sum-squares obtained by moving from the A-level model to the B-level model. Symbolically, \( R(\beta|\mu, \alpha) = R(\alpha, \beta|\mu) - R(\alpha|\mu) \).

Does the ANOVA presented in Table 3 mean that 25.4% of the sum-squares are “due to” the A-level (\( \alpha \)) effects and 64.8% “due to” the B-level (\( \beta \)) effects? No. It means that 25.4% of the variance is explained by placing the raw observations in A-level categories without controls for lower-level effects.
The problem flows from the fact that the model is nested. Recall from Observation 3–Observation 4 that one cannot estimate the pure A-level effects. In fact, fitting the A-level effects amounts to estimating the A1 dummy variable, and from (5) we see that this picks up a mixture of A-level and B-level effects. Just as one cannot estimate the pure $\alpha$’s using regression, one cannot deduce their contribution to the sum-squares from the $R^2$ accompanying $R(\alpha|\mu)$.

### 3 Variance Components

In order to measure the relative impact of the A-level and B-level effects on sum-squares (or on total variance) we must turn to the estimation of variance components. This requires a reinterpretation of the basic model (1). It is often said that the new view is that the $\alpha_i$ and the $\beta_{ij}$ are considered to be random rather than fixed effects. But this does not really capture the essence of the reinterpretation. The values of these “effects” are always fixed for any given data set and they are always unobservable. The new interpretation concerns the process by which these effects were generated. In a random components model we assume that the $\alpha_i$ are independent draws from a distribution with mean zero and variance $\sigma^2_\alpha$. Similarly, the $\beta_{ij}$ are independent draws from a different, independent, distribution with mean zero and variance $\sigma^2_\beta$. Examining (1) in this light, it should be clear that

$$
\sigma^2_y = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\epsilon.
$$

The new estimation problem is to estimate $\mu$, and the three variance components $\sigma^2_\alpha$, $\sigma^2_\beta$, and $\sigma^2_\epsilon$. To see how this can be done, consider the expected value of a single observation:

$$
E[y^2_{ijk}] = E[\mu^2] + 2E[\mu \alpha_i] + 2E[\mu \beta_{ij}] + 2E[\mu \epsilon_{ijk}]
$$
\[ + \ E[\alpha_i^2] + 2E[\alpha_i\beta_{ij}] + 2E[\alpha_i\epsilon_{ijk}] \]
\[ + \ E[\beta_{ij}^2] + 2E[\beta_{ij}\epsilon_{ijk}] \]
\[ + \ E[\epsilon_{ijk}^2] \]

Because the distributions of the effects have mean zero, and because of independence, only the first term in each line has a non-zero expected value. Hence,

\[ E[y_{ijk}^2] = \mu^2 + \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\epsilon. \] (6)

Similarly, the expected value of any cross product \( y_{ijk} \cdot y_{pqr} \) is a linear combination of \( \mu^2 \) and the variance components. More generally, the expected value of any quadratic form in the observations is a linear combination of \( \mu^2 \) and the variance components. Thus, one approach to estimation is to pick four quadratic forms and equate the expected values to the observed values—then four equations in four unknowns can be solved. The standard choice is quadratic forms used in ANOVA: \( T_0, T_m, T_A, \) and \( T_{AB} \).

\[
E(T_0) = E\left[ \sum_{i,j,k} y_{ijk}^2 \right] = N\mu^2 + N\sigma^2_\alpha + N\sigma^2_\beta + N\sigma^2_\epsilon \] (7)

\[
E(T_m) = E\left[ \frac{y_{i..}^2}{N} \right] = N\mu^2 + \frac{1}{N} \sum_i c^2_i \sigma^2_\alpha + \frac{1}{N} \sum_{ij} c^2_{ij} \sigma^2_\beta + \sigma^2_\epsilon \] (8)

\[
E(T_A) = E\left[ \sum_i \frac{y_{i..}^2}{c_i} \right] = N\mu^2 + N\sigma^2_\alpha + \sum_i \left[ \frac{\sum_j c^2_{ij}}{c_{i.}} \right] \sigma^2_\beta + a\sigma^2_\epsilon \] (9)

\[
E(T_{AB}) = E\left[ \sum_{ij} \frac{y_{ij}^2}{c_{ij}} \right] = N\mu^2 + N\sigma^2_\alpha + N\sigma^2_\beta + b\sigma^2_\epsilon \] (10)

Plugging in the numbers from our example problem, the equations to be solved are:

\[
\begin{align*}
11\mu^2 &+ 11 \sigma^2_\alpha &+ 11 \sigma^2_\beta &+ 11 \sigma^2_\epsilon &= 8628.00 \\
11\mu^2 &+ 5.55 \sigma^2_\alpha &+ 2.27 \sigma^2_\beta &+ 1 \sigma^2_\epsilon &= 4241.46 \\
11\mu^2 &+ 11 \sigma^2_\alpha &+ 4.60 \sigma^2_\beta &+ 2 \sigma^2_\epsilon &= 5357.97 \\
11\mu^2 &+ 11 \sigma^2_\alpha &+ 11 \sigma^2_\beta &+ 5 \sigma^2_\epsilon &= 8200.50
\end{align*}
\]

Solving for the unknowns gives \( \hat{\mu}^2 = 286.0 \) and the variance components shown in Table 4. Manipulating the equations in the unknowns produces the table of expected values of sum-squares shown in Table 5. This table helps us understand the relationships between variance components and the results of ANOVA regression. In particular, note that there is no clean correspondence between these components of sum-squares
Table 4: Variance Components Estimation

<table>
<thead>
<tr>
<th>Component</th>
<th>Estimate</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_\alpha$</td>
<td>16.4</td>
<td>0.033</td>
</tr>
<tr>
<td>$\sigma^2_\beta$</td>
<td>410.7</td>
<td>0.824</td>
</tr>
<tr>
<td>$\sigma^2_\epsilon$</td>
<td>71.3</td>
<td>0.143</td>
</tr>
<tr>
<td>Total</td>
<td>498.38</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 5: Expected Values of sum-squares

<table>
<thead>
<tr>
<th>Source</th>
<th>sum-squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R(\mu)$</td>
<td>$11\mu^2 + 5.55\sigma^2_\alpha + 2.27\sigma^2_\beta + \sigma^2_\epsilon$</td>
</tr>
<tr>
<td>$R(\alpha</td>
<td>\mu)$</td>
</tr>
<tr>
<td>$R(\beta</td>
<td>\mu, \alpha)$</td>
</tr>
<tr>
<td>$R(\alpha, \beta</td>
<td>\mu)$</td>
</tr>
<tr>
<td>$SSE$</td>
<td>$6\sigma^2_\epsilon$</td>
</tr>
<tr>
<td>$SST_m$</td>
<td>$5.45\sigma^2_\alpha + 8.73\sigma^2_\beta + 10\sigma^2_\epsilon$</td>
</tr>
<tr>
<td>$SST$</td>
<td>$11\mu^2 + 11\sigma^2_\alpha + 11\sigma^2_\beta + 11\sigma^2_\epsilon$</td>
</tr>
</tbody>
</table>

and the variance components. Examine the line for $SST_m$; it shows $E[SST_m] = 5.45\sigma^2_\alpha + 8.73\sigma^2_\beta + 10\sigma^2_\epsilon$. This is 10 times the estimated error variance ($\hat{\sigma}^2_\epsilon$) and is the denominator of $R^2$. Define the components of $SST_m$ as $S_\alpha = 5.45\sigma^2_\alpha$, $S_\beta = 8.73\sigma^2_\beta$, and $S_\epsilon = 10\sigma^2_\epsilon$; each component clearly corresponds to one of the variance components. Now,

$$R^2(\alpha|\mu) = \frac{S_\alpha + 0.27S_\beta + 0.1S_\epsilon}{S_\alpha + S_\beta + S_\epsilon}$$  \hspace{1cm} (11)$$

$$R^2(\beta|\alpha, \mu) = \frac{0.73S_\beta + 0.3S_\epsilon}{S_\alpha + S_\beta + S_\epsilon},$$  \hspace{1cm} (12)$$

and

$$SSE/SST_m = \frac{0.6S_\epsilon}{S_\alpha + S_\beta + S_\epsilon}$$  \hspace{1cm} (13)$$

First, note that whereas the “true” fraction of error in the total sum-squares (variance) is $S_\epsilon/(S_\alpha + S_\beta + S_\epsilon)$, (13) shows that $SSE/SST_m$ is only expected to be 6/10 of this amount. Furthermore, the incremental $R^2$ due to fitting $\beta$ after $\alpha$ and $\mu$ depends on both $\sigma^2_\beta$ and $\sigma^2_\epsilon$. 

11
This problem arises because the coefficients of $\sigma^2_\beta$ and $\sigma^2_\epsilon$ in (9) are not zero. Consequently, there are contributions from the B-level effects and from error effects in $R^2(\alpha|\mu)!$

To examine this issue more closely, recall that $R^2(\alpha|\mu)$ is the ratio of $T_\alpha - T_\mu$ to $T_\alpha - T_\mu$. Normalize by dividing both numerator and denominator by $N$ and let $Q_\beta$ be the resulting coefficient of $\sigma^2_\beta$ in the numerator. Then

$$Q_\beta = \sum_{i=1}^{a} \left[ \frac{c_i \cdot}{N} \right] \sum_{j=1}^{b_i} \left[ \frac{c_{ij}}{c_i \cdot} \right]^2 - \sum_{i=1}^{a} \sum_{j=1}^{b_i} \left[ \frac{c_{ij}}{N} \right]^2. \quad (14)$$

Some intuition about the size of $Q_\beta$ may be found from the case for balanced data. With balanced data, $b_i = b$ for all $i$ and $c_{ij} = c$ in each cell. Then $N = abc$ and $c_i \cdot = bc$. Working with (14) gives

$$Q_\beta = \frac{1}{b} - \frac{1}{ab}. \quad (15)$$

This clearly does not disappear as the sample size increases, but does get smaller as the number of B-level classes increases (e.g., business-units within corporations). However, in the case of unbalanced data, the value of $Q_\beta$ increases not only with average number of B-level classes within each A-level class, but also with the degree to which observations are “concentrated” within both A-level and B-level classes. To be more precise, recall that the Herfindahl index is $h = \sum^n s_i^2$ where $s_i$ is the fraction or share of a quantity in the $i$'th category ($s_\cdot = 1$). Some algebra easily shows that $h = (1 + \sigma^2_s/s^2)/n$: $h$ falls in the number $n$ of classes and rises with the relative variance of inter-class shares. Now define $H = \sum_{ij} c_{ij}^2/N^2$, the Herfindahl index for observations taken across all of the B-level categories, and define $h_i = \sum_j c_{ij}^2/c_{i \cdot}^2$, the Herfindahl index for observations across the B-level categories within the $i$'th A-level group. Now (14) can be rewritten as

$$Q_\beta = \sum_i \frac{c_i \cdot}{N} h_i - H. \quad (16)$$

In this expression the Herfindahl indices measure the concentration of observations across B-level units; the $h_i$ are the concentrations within each A-level group, and $H$ is the overall concentration. Thus, $Q_\beta$ is the difference between the weighted average of the $h_i$’s and $H$. Consequently, the more “unbalanced” the data—the more some corporations have more business-units than others—the larger will be $Q_\beta$. And as $Q_\beta$ increases, so does the contribution of $\sigma^2_\beta$ to $R^2(\alpha|\mu)$. The $R^2$ attached to the entry of the A-level effects overstates their importance—the overstatement increases with the variance among the B-level effects and increases as the data structure becomes less balanced.
Table 6: Variance Components Estimates: Restricted Model (Year Effects and $C_{\alpha\beta}$ Removed)

(Table 3 in Rumelt[1991])

<table>
<thead>
<tr>
<th>Component</th>
<th>Sample A</th>
<th>Sample B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std. Est. Error</td>
<td>Std. Est. Error</td>
</tr>
<tr>
<td>Industry-Year</td>
<td>21.92 2.04</td>
<td>22.09 2.31</td>
</tr>
<tr>
<td>Industry</td>
<td>23.26 4.72</td>
<td>16.55 4.26</td>
</tr>
<tr>
<td>Corporation</td>
<td>2.25 3.84</td>
<td>6.74 3.31</td>
</tr>
<tr>
<td>Business-unit</td>
<td>129.63 6.91</td>
<td>181.50 7.04</td>
</tr>
<tr>
<td>Error</td>
<td>102.51 2.18</td>
<td>184.06 3.04</td>
</tr>
<tr>
<td>Total</td>
<td>279.56 100.00</td>
<td>410.95 100.00</td>
</tr>
</tbody>
</table>

3.1 A Re-Look at the LOB Data

The result in Rumelt [1991] which generated the most interest and controversy was the “zero” corporate effect for Sample A. Table 6 reproduces “Table 3” from that paper and Table 7 reproduces the original “Table 2.” As can be seen, the estimated corporate effects is very small. By contrast, the second half of Table 7 shows that the $R^2$ due to corporate effects was 17.6%.\(^1\) How can these results be reconciled?

Working with the original data, dropping the insignificant year effects, and adopting the symbols used in the original paper, gives

\[
SST_m = 1,935,767.42 \\
\mathbb{E}[SST_m] = 6894.53\sigma^2_\alpha + 6903.89\sigma^2_\beta + 6922.61\sigma^2_\delta + 6928.07\sigma^2_\phi + 6931\sigma^2_\epsilon \tag{17} \\
T_B - T_m = 340,475.22 \\
\mathbb{E}[T_B - T_m] = 1727.17\sigma^2_\alpha + 6903.89\sigma^2_\beta + 447.61\sigma^2_\delta + 1760.72\sigma^2_\phi + 456\sigma^2_\epsilon \tag{18}
\]

Again adopting the convention that $S_\alpha$ represents the “amount” of the $\sigma^2_\alpha$ effects in $SST_M$, we have $SST_m = S_\alpha + S_\beta + S_\delta + S_\phi + S_\epsilon$. Numerically,

\[
S_\alpha = 6894.53 \cdot \sigma^2_\alpha = 6894.53 \cdot 23.26
\]

\(^1\)Because the impact of the year-effects was so small the corporate effects, which actually entered after year effects, can be considered to have entered first to a very good approximation.
Table 7: Fixed-Effects ANOVA Results
(Second Half of Table 2 in Rumelt[1991])

<table>
<thead>
<tr>
<th>Source</th>
<th>Sample A</th>
<th>Sample B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>Incr. R²</td>
</tr>
<tr>
<td>Year</td>
<td>3</td>
<td>0.0003</td>
</tr>
<tr>
<td>Corporation</td>
<td>456</td>
<td>0.176</td>
</tr>
<tr>
<td>Industry</td>
<td>241</td>
<td>0.153</td>
</tr>
<tr>
<td>Business-unit</td>
<td>1076</td>
<td>0.340</td>
</tr>
<tr>
<td>Industry-Year</td>
<td>721</td>
<td>0.096</td>
</tr>
<tr>
<td>Total Model</td>
<td>2497</td>
<td>0.765</td>
</tr>
<tr>
<td>Error</td>
<td>4434</td>
<td>0.235</td>
</tr>
<tr>
<td>Total</td>
<td>6931</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

aSignificant at 0.0001 level.
bSignificant at 0.0005 level.

Thus, the “true” fraction of variance due to corporate effects is $S_\beta / SST_m = 0.0008$. Working with these definitions, we have

$$R^2(\beta|\mu) = \frac{T_B - T_m}{SST_m} = 0.176.$$  

This fraction is much larger than 0.0008 because there are elements in the numerator of this $R^2$ which are not due to $\sigma_\beta^2$. Taking expected values of the numerator and denominator, we have

$$\frac{E[T_B - T_m]}{E[SST_m]} = \frac{0.251S_\alpha + 1.0S_\beta + 0.065S_\delta + 0.254S_\phi + 0.066S_\epsilon}{S_\alpha + S_\beta + S_\delta + S_\phi + S_\epsilon} = 0.176$$  \hspace{1cm} (19)

Going further, by entering the numerical values for $S_\alpha, S_\beta, etc.$, we can decompose $R^2(\beta|\mu)$ as shown in Table 8.
Table 8: Components of $R^2(\beta|\mu)$

<table>
<thead>
<tr>
<th>Source</th>
<th>Portion of $R^2$</th>
<th>Percentage of $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry ($\sigma^2_{\alpha}$)</td>
<td>2.07</td>
<td>11.8</td>
</tr>
<tr>
<td>Corporate ($\sigma^2_{\beta}$)</td>
<td>0.80</td>
<td>4.6</td>
</tr>
<tr>
<td>Industry-Year ($\sigma^2_{\delta}$)</td>
<td>0.51</td>
<td>2.9</td>
</tr>
<tr>
<td>Business-Unit ($\sigma^2_{\phi}$)</td>
<td>11.79</td>
<td>67.0</td>
</tr>
<tr>
<td>Error ($\sigma^2$)</td>
<td>2.41</td>
<td>13.7</td>
</tr>
<tr>
<td>Total</td>
<td>17.59</td>
<td>100.0</td>
</tr>
</tbody>
</table>

As can be seen, the majority of $R^2(\beta|\mu)$ is not due to corporate effects at all; it actually flows from business-unit effects. Corporate effects account for only 4.6% of the value of $R^2(\beta|\mu)$. This is partly due to the fact that the estimated $\sigma^2_{\beta}$ is small. However, it is also a direct result of the nested structure of the model. To demonstrate this, suppose that the estimation of corporate effects had risen to equal the industry effects ($\sigma^2_{\beta} = \sigma^2_{\alpha} = 23.26$) and that the business-unit effects were, consequently, estimated to be smaller ($\sigma^2_{\phi} = 108.64$). Plugging these new estimates into expressions for expected sum-squares gives an expected value of $R^2(\beta|\mu) = 23.18$. However, the data structure ensures that even with these changes, the largest contributor to $R^2(\beta|\mu)$ are still the (reduced!) business-unit effects: they would account for 43% of the $R^2$, with corporate and industry effects accounting for 36% and 9%, respectively. The $R^2$ from ANOVA is simply not a reliable indicator of relative importance.

References


