Making Use of Atypical Regression Models for Theory Building

Ernest C. Davenport, Jr., PhD[1], Haijiang Kuang[2], PhD, Mark L. Davison, PhD[1],
Kyle Nickodem[1], Qinjun Wang[1]
[1]University of Minnesota, [2]Pearson Corporation

Abstract

While investigators typically use one regression model to ascertain information regarding a theory, this presentation gives six equivalent regression models (for a 2-predictor model) with predictors and/or their mutual residuals as predictors that can lead to richer theory building. While each model is equivalent in significance and variance accounted for, the mix of raw and residual score predictors provide for testing different aspects of a theory. We have named the approach Total Information Regression Analysis (TIRA). We demonstrate the approach with data from the 2001 edition of Baccalaureate and Beyond. Undergraduate GPA is the criterion. For simplicity we use only two predictors, SATM and SATV (and their errors), but the approach is readily adaptable to more predictors (which will increase the number of potential equivalent models). We also give a geometric rationale for our procedure. The geometric rationale is especially helpful as half of the models presented are not obviously equivalent.

Keywords: Regression, Geometry, Models, Theory-Building

1. Introduction

This paper utilizes notation and ideas from Schey (1993) to present a geometric representation of regression analysis that demonstrates a counter-intuitive finding relative to prediction from error terms. We can use this finding to develop additional regression models that provide for better understanding of a phenomenon. Below we first present three well understood regression models, then the rationale for a counter-intuitive model. This rationale will be extended to include other counter-intuitive models. We next provide an example using these models to do a "Total Information Regression Analysis". The paper ends with extensions of the procedure and comments relative to theory building.

It is well known that the following three models are equivalent.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad \text{(Model 1)}$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 U_2 + \varepsilon \qquad \text{(Model 2)}$$
$$Y = \beta_0 + \beta_1 U_1 + \beta_2 X_2 + \varepsilon \qquad \text{(Model 3)}$$

Where U_2 represents the unique portion of X_2 above and beyond that which can be predicted from X_1 and U_1 represents the unique portion of X_1 above and beyond that which can be predicted from X_2. The last two components just provide the incremental prediction of the second variable given that the first is already in the model; essentially the Type 3 Sums of Squares (Maxwell and Delaney, 1990). Note also that U_1 is orthogonal to X_2 as these two variables are unrelated and U_2 is orthogonal to X_1 as these two variables are also unrelated.

What is not as intuitive is that

$$Y = \beta_0 + \beta_1 U_1 + \beta_2 U_2 + \varepsilon \qquad \text{(Model 4)}$$

is also equivalent to the other three models. The reason for this seeming dilemma is that the predictors in Model 4 contain only residual terms which are the unique portions of the predictors above and beyond that of the other predictor. U_1 can be interpreted as all that is in X_1 that is not related to X_2. Similarly, U_2 can be interpreted as all that is in X_2 that is not related to X_1. Therefore, it would appear that the common part of X_1 and X_2 is missing as a part of the prediction equation from Model 4.

The geometric approach that we are promoting allows investigators to "see" regression. It provides a mechanism by which one can view the vectors and projections which underlie regression analysis. If a picture is worth a thousand words, this approach should be useful in conveying a breadth of information in a concise manner. Thus, in addition to merely calculating regression coefficients, one can also have a geometrical picture of the effect of regressing each predictor, alone and/or simultaneously. As stated above, each of the models (1 - 4) are equivalent and Schey's (1993) geometric approach provides an easy technique to see the equivalence. It also provides an easy method to prove various relationships between the independent variables, the unique portions of the independent variables, and the dependent variable.

2. Specifics on the Models

For the four models given above, Y is the dependent variable. β_0 is the constant term in the regression equation. β_1 and β_2 are the regression coefficients applied to the first and second predictors in each model, respectively. Note that the estimates for β_0, β_1, and β_2 differ for each model since the predictors are different for each model. Still, each model is equivalent with respect to amount of variance accounted for and each leads to the same predicted value, Y^*, for each observation. Thus, these models have the same SS_E, $\Sigma(Y - Y^*)^2$, same significance, same degrees of freedom, and same R^2.

X_1 represents the first predictor (i.e. independent variable) and X_2 represents the second. These are the original variables. U_1 and U_2 are related to the original variables. U_1 is the unique portion of X_1 orthogonal to X_2. It is derived by obtaining the errors in predicting X_1 from X_2. $U_1 = X_1 - X_1^*$ where X_1^* is the predicted value of X_1 if you use X_2 as the predictor, $X_1^* = b_0 + b_1 X_2$. Correspondingly, U_2 is the unique portion of X_2 orthogonal to X_1. It is derived by obtaining the errors in predicting X_2 from X_1. $U_2 = X_2 - X_2^*$ where X_2^* is the predicted value of X_2 if you use X_1 as the predictor (e.g. $X_2^* = b_0 + b_1 X_1$). Finally, ε is the error term for each model. Note that the ε_i values (error for observation i) will be the same for each model.

It is fairly well established that the first three models are equivalent. Model 1 uses X_1 and X_2 to predict Y. Model 2 uses X_1 and the unique part of X_2 which is orthogonal to X_1 to predict Y. Finally, Model 3 uses X_2 and the unique part of X_1 which is orthogonal to X_2 to predict Y. Given that the common part of X_1 and X_2 that is related to Y occurs in both X_1 and X_2, it is intuitive that adding U_2 to predict above and beyond X_1 (Model 2) and adding U_1 to predict above and beyond X_2 (Model 3) should produce models equivalent to Model 1.

In contrast, the equivalence of the fourth model to the other three is not as intuitive. Model 4 takes only the unique portions of X_1 and X_2 to predict Y which seems to ignore the prediction capability inherent in the common portion of these variables. Thus, only in the rare case where X_1 and X_2 are orthogonal would one believe this model to be equivalent to the others. The fallacy in this logic is easily exposed by simple geometric principles. From geometry we know that any two vectors in a two-dimensional plane will span the plane. Moreover, we know that U_1 is orthogonal to X_2 and that its orthogonal projection is still in the plane spanned by X_1 and X_2. Likewise, U_2 is orthogonal to X_1, but is also remains in the plane spanned by X_1 and X_2. These revelations lead to the possibility of two new equivalent models that are even more counter-intuitive:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 U_1 + \varepsilon \qquad \text{(Model 5)}$$

$$Y = \beta_0 + \beta_1 X_2 + \beta_2 U_2 + \varepsilon \qquad \text{(Model 6)}$$

Note that Model 5 seems to be a result of X_1 alone. It is devoid of X_2. Remember that U_1 is orthogonal to X_2. It is the error remaining in X_1 after predicting X_1 with X_2. Thus, by definition U_1 is unrelated to X_2. Model 6 appears to be devoid of X_1. It is composed of X_2 and U_2; where by definition U_2 is unrelated to X_1.

3. Example

The data for our example come from the 2001 edition of Baccalaureate and Beyond[1] (B&B). B&B is one of the national surveys from the National Center for Education Statistics (NCES) to provide information on education in the U.S. This particular set of data examines students' education and work experiences after they complete a bachelor's degree, with special emphasis on the experiences of new elementary and secondary teachers. Note that we only took students with majors in the STEM disciplines of physical science, mathematics, computer information science, and engineering. Note, too, that we only took students with SAT[2] scores as our predictors are SATM (the quantitative part of the test) and SATV (the critical reading part of the test). Note that the student's overall grade point average (GPA) is the criterion. Students in STEM disciplines with complete data for the SAT and GPA led to a sample of approximately 525 students. All of our analyses were appropriately weighted and we use a design effect of two to adjust our significance tests (essentially cutting the degrees of freedom for error in half before obtaining the MSE used in the significance tests) for any lack of independence of the survey participants..

Preliminary and summary information are shown in Tables 1 – 3. Table 1 shows that separately SATM and SATV are highly significant predictors of GPA ($P < 0.001$); indicating that the probability of the relationship being a chance event is essentially zero. Thus, we conclude that SATM and SATV are both significantly related to the GPA of students with college majors in STEM disciplines. Given that these are students with STEM majors, it is not surprising that SATM accounts for more variance in GPA than SATV. Table 2 is the Analysis of Variance table for the regression of the two-predictor models (1-6) predicting GPA. Given that we have said that Models 1 – 6 are equivalent, the overall results for each of the models will be the same (see Table 2). The two-predictor regression model is significant suggesting that the predictors add significantly to the prediction of the criterion. Therefore, SATM and SATV simultaneously are significantly related to GPA for students with STEM majors. Note that the adjusted degrees of freedom for error has essentially been cut in half to account for possible dependencies in the data. Finally, note that the proportion of variance accounted for by the two-predictor model, 0.1178, is not much different than the amount of variance accounted for by SATM alone (see Table 1). This suggests little additional predictability for adding SATV to the model over and beyond that of SATM.

Table 1 - Correlation of SAT Scores with GPA

	Corr	T	Pvalue	R^2
SATM	0.340	5.878	0.000	0.115
SATV	0.230	3.839	0.000	0.053

[1] https://nces.ed.gov/surveys/b&b/
[2] https://sat.collegeboard.org/scores

Table 2 – Analysis of Variance for Models 1 - 6

	Sums of Squares	Df	Mean Square	F	P_Value
Model	18131282	2	9065641	17.685	0
Error	135842929	265	512614.8		
Total	153974212				
R^2	0.1178				

Table 3 – Correlations and Angular Separation of Variables

	GPA	SATM	SATV	P_GPA	E_SATM	E_SATV
GPA	1.000	0.340	0.230	0.343	0.255	0.049
	0.0	70.1	76.7	69.9	75.2	87.2
SATM	0.340	1.000	0.555	0.990	0.832	0.000
	70.1	0.0	56.3	8.3	33.7	90.0
SATV	0.230	0.555	1.000	0.669	0.000	0.832
	76.7	56.3	0.0	48.0	90.0	33.7
P_GPA	0.343	0.990	0.669	1.000	0.743	0.144
	69.9	8.3	48.0	0.0	42.0	81.7
E_SATM	0.255	0.832	0.000	0.743	1.000	-0.555
	75.2	33.7	90.0	42.0	0.0	123.7
E_SATV	0.049	0.000	0.832	0.144	-0.555	1.000
	87.2	90.0	33.7	81.7	123.7	0.0

1st line correlation / 2nd line is the angular separation

All correlations are significant ($P < 0.001$) except GPA and E_SATV (n.s.) and SATM versus E_SATV and SATV and E_SATM which are known orthogonal

Table 3 shows the inter-correlations of all of the variables we use in the analyses. GPA, SATM, and SATV were from the data and discussed above. P_GPA is the predicted value of GPA for any of the two predictor models (1 – 6). Given the equivalence of the models the predicted value of each will be the same. From Table 3 we know that SATV and SATM are correlated (0.555). This is as expected given that although the scores represent different skill sets, positive manifold in intelligence suggests that in general students with high SATM scores will be the same students with high SATV scores (Hakstian and Cattell, 1978). E_SATM is the error with using SATV to predict SATM. The common part that SATV shares with SATM as shown by their correlation has been partialled out of SATM resulting in E_SATM. In practical terms E_SATM is the portion of SATM that is unrelated to SATV. It is the mathematics portion that is devoid of any relationship to the critical reading (Verbal) portion. Mathematically, E_SATM is orthogonal to SATV. Similarly, E_SATV is the error using SATM to predict SATV and thus it is the verbal portion devoid of the relationship to mathematics. Notice that all of the correlations are significant at $P < 0.001$ with the exception of GPA and E_SATV where there is a non-significant relationship and the variables that are known to be orthogonal (SATM versus E_SATV and SATV versus E_SATM). Thus, while SATV is significantly correlated with GPA, when the common portion of SATV that is shared by SATM is removed, the remaining part of SATV (E_SATV) is no longer predictive of GPA for STEM majors. The last notable feature of Table 3 is that the results are given in line pairs. The first line of a pair holds the correlations of the variables while the second line of each pair contains the angular separation of the variables (vectors) of interest. For example, the correlation of SATV and SATM is 0.555 and the vectors represented by the scores of SATV and SATM have an angular separation of 56.3 degrees. This result is due to the correlation between two variables being equal to the cosine of the angle between their vectors (Schey, 1993). Thus, $COS^{-1}(0.555) = 56.289°$.

Table 4 shows the results for the T tests of the estimators in Model 1. Note that SATM is significant and SATV is not. This T test is a Type 3 test that tests the significance of the predictor in question as to whether it makes a unique contribution given that the other variable is already in the model. Thus, SATM is still a significant predictor above and beyond SATV. However, SATV adds little to the predictability if SATM is already in the model. Note that this is not surprising given that our sample is comprised of students from STEM majors which may relate more to the skill set measured by SATM.

Table 4 - Model 1: $GPA = \beta_0 + \beta_1 SATM + \beta_2 SATV + \varepsilon$ $R^2 = 0.1178$

	Stand_Est	Df_adj	SSM	T	df	P	Unique R^2
SATM	0.3067	265	10020962	4.421	265	0.000	0.065
SATV	0.0593	265	374193	0.854	265	0.394	0.002

Figure 1 shows the geometric representation for the space spanned by the predictors in Model 1. This and subsequent figures are included to show the equivalence of each of the models in their prediction of GPA. Note that the Y vector (GPA) also starts at point (0,0) on this figure but is in a higher dimensional space and projects down into the space of the predictors (represented by Figure 1) at three points. The first is at the endpoint of the vector labeled SATV. SATV is the prediction vector for GPA in the space spanned by SATV and SATM if only SATV is used as a predictor. It equals the square root of the sums of squares for the model when only SATV is used as the predictor ($\sqrt{374193} = 611.7$). The vector that projects down from the criterion vector is the error vector for this model. Given that we are using Least Squares regression, we know that this error is minimized and thus the projection is orthogonal to the SATV / SATM plane at SATV (an orthogonal projection makes the error vector for this prediction as short as possible). Moreover, the length of this error vector is just the square root of the sums of squares error for this model with only SATV as the predictor.

We get a similar error vector when we use SATM as the sole predictor. Then SATM becomes the predicted vector for the criterion in the SATV / SATM plane. The length of this vector is the square root of the sums of squares model when SATM is the only predictor. The new error will be the orthogonal projection from the criterion vector to SATM. Finally, when both predictors are used simultaneously (SATV and SATM), the criterion vector projects into the spaced spanned by SATV / SATM at Y* with the subsequent error for the 2-predictor model being an orthogonal projection from the GPA vector to this point. What we will show in the subsequent figures is that regardless of the predictors used for Models 1 – 6, Y* will always be in the same place and thus its predicted values (Y*), Sums of Squares for the Model, and Sums of Squares for Error will always be the same.

There are two vectors in Figure 1 that have yet to be explained, M|V and V|M. M|V is the additional sums of squares necessary to go from just having SATV as the predictor to having both SATV and SATM as predictors. It can be found by obtaining the Type 3 Sums of Squares for adding SATM to a model that has SATV in it. The length of M|V is just the square root of these sums of squares. It is the additional predictability that is obtained by adding SATM to a model with SATV. Note too that this vector is orthogonal to SATV. Thus, by Pythagorean's theorem the Sums of Squares for SATV and the unique sums of squares for SATM given SATV is already in the model equals the Sums of Squares for Y* (the squared length of Y*) given, the 2-predictor model. Note that V|M can be described similarly. It is the amount necessary to add to SATM to get to Y*. We already know that this is a small quantity as SATV did not add much given that SATM was already in the model.

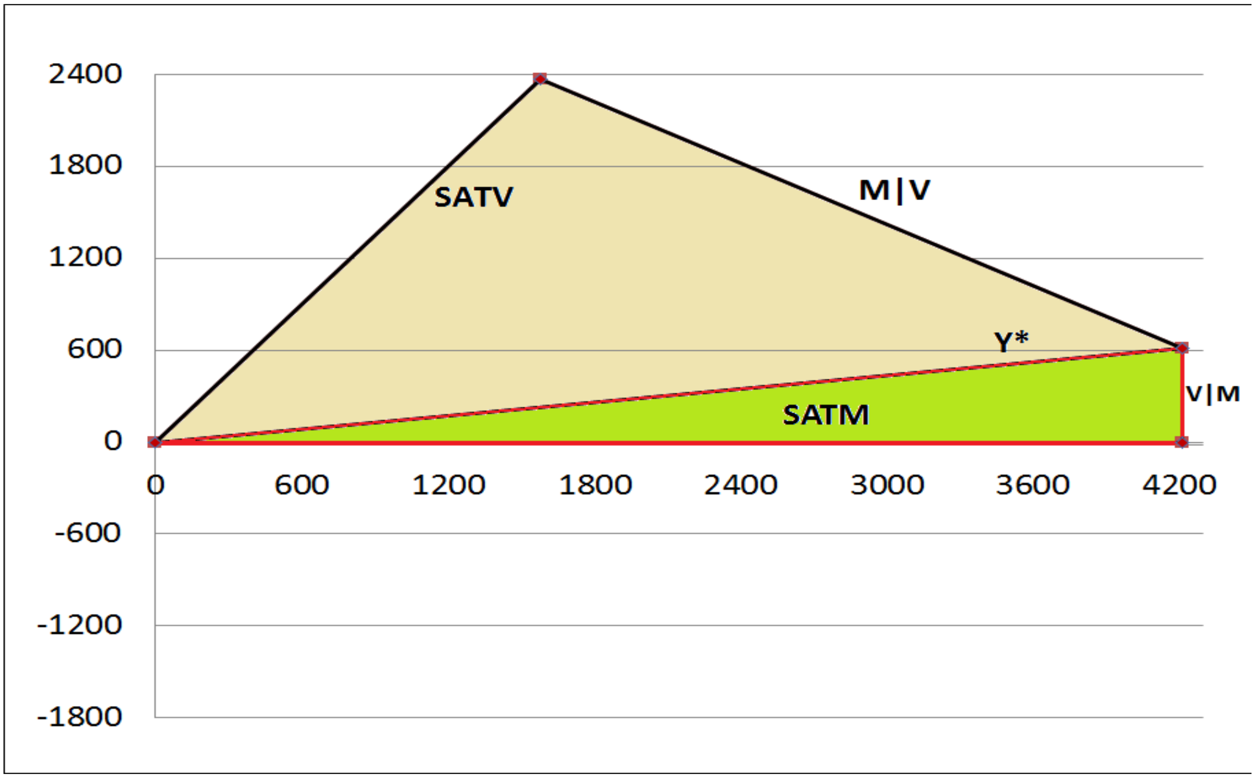

Figure 1: Geometrical Representation of Model 1

Table 5 shows the results of Model 2. Note that the predictors in this model are SATM and E_SATV (the unique part of SATV above and beyond SATM). Again, the T tests are the test of significance given that the other predictor is in the model. The difference in the model this time is that SATM and E_SATV are unrelated as E_SATV is what remains of SATV after it has been cleansed of its relationship with SATM. Thus, the unique predictability for SATM for this model is the same as the original correlation in Table 1, 0.115 (the significance is different because of the influence of E_SATV on the full model here). On the other hand given that E_SATV is cleansed of SATM, it represents the unique contribution of SATV given SATM is already in the model. Thus, E_SATV accounts for the same sums of squares (and significance) as the Type 3 test for SATV above (see Table 4). Again, we see that SATM is predictive, but the unique contribution of SATV (E_SATV) above and beyond SATM is not.

Table 5: Model 2: $Y = \beta_0 + \beta_1 SATM + \beta_2 E\_SATV + \varepsilon$ **$R^2 = 0.1178$**

	Stand_Est	Df_adj	SSM	T	df	P	Unique R^2
SATM	0.3396	265	17757089	5.886	265	0.000	0.115
E_SATV	0.0493	265	374193	0.854	265	0.394	0.002

From Figure 2 we see that SATM and Y* are still in the same place as in Figure 1. Note, again, that SATM is the vector predicting the criterion in the plane spanned by SATM / E_SATV that is the orthogonal projection of the criterion vector onto that vector solely. Again, Y* is the projection of the criterion vector into the space spanned by SATM / E_SATV when using both vectors as predictors. E_SATV is the projection of the criterion vector onto the single predictor vector, E_SATV. Note that the vector E_SATV is perpendicular to the vector represented by SATM, demonstrating the effect of SATM being partialled out of SATV. Also note that given Pythagorean's theorem that the sums of squares of the predictors separately are additive to that for the predictors simultaneously. Thus, what is needed to go from predicting solely with SATM to the results (Y*) when predicting with both SATM and E_SATV, EV|M, simply equals E_SATV. Likewise what is needed to go from predicting with E_SATV to predicting with both E_SATV and MATH, M|EV, simply equals SATM. This picture clearly shows that length of E_SATV equals EV|M. Note that this model is just as expected – we take the first predictor and then the second cleansed of the effect of the first and we get an additive model whose predictability equals that of the two original variables.

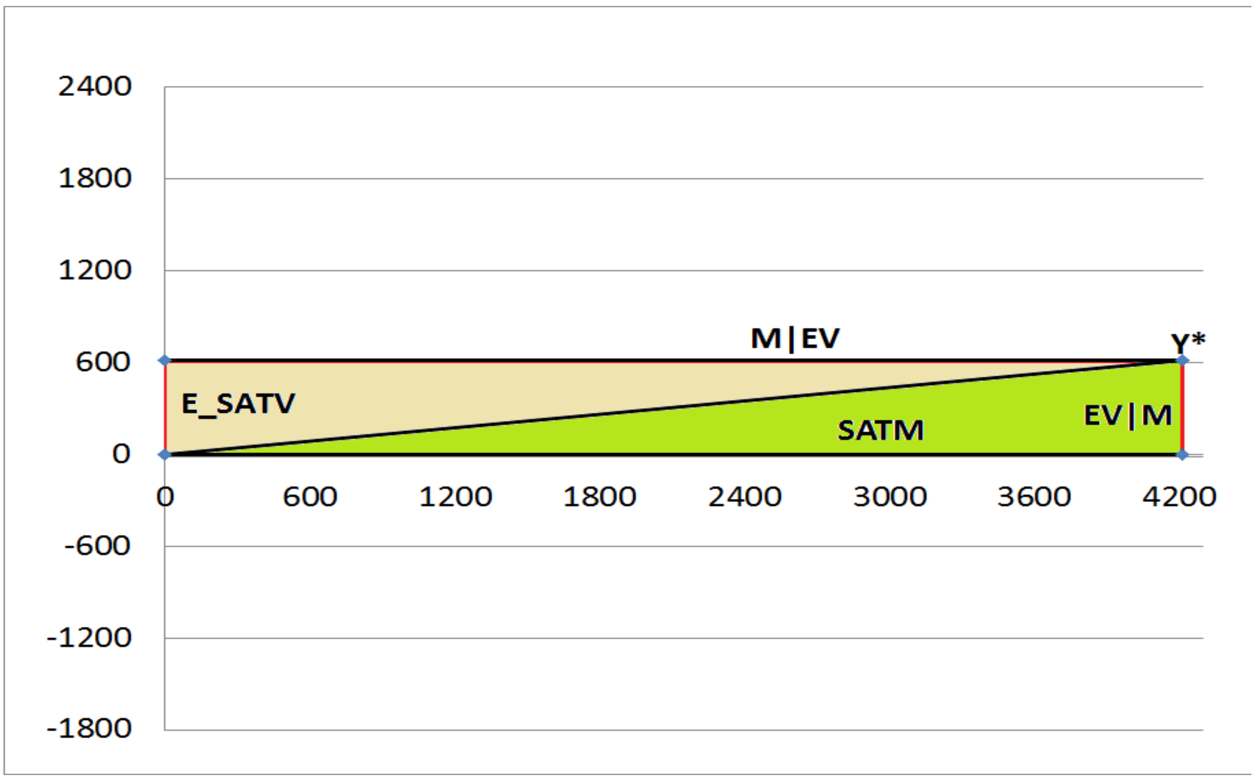

Figure 2: Geometrical Representation of Model 2

Table 6 holds the results for Model 3. Note that Model 3 is just the other side of the coin from Model 2. In this case we maintain the second predictor, SATV, and take the first, SATM, cleansed of the relationship of the second to result in E_SATM. This model is also fairly intuitive. The predictability of SATV should be the same as represented in its original correlation and E_SATM should have the same predictability as its unique portion from Model 1. Again, we expect our predictors to be orthogonal and our sums of squares for the 1-predictor models to be additive to that for the 2-predictor model. Note here that SATV is significant. This result tells us that there is still a verbal component that is predictive of GPA even for STEM majors. However, this verbal component is confounded with the quantitative test. There is a verbal component to the quantitative test that is assigned to the quantitative test when it is in the model. Only when the model is cleansed of this influence can this part of the test make itself known separately.

Figure 3 is as expected. The prediction vector of E_SATM is perpendicular to that of SATV. The vectors SATV and Y* are still in the same place as in Figure 1 and still represent the same projections. The squared lengths of the vectors that correspond to their sums of squares is additive. Finally, S_SATM equals EM|V which is the additional amount needed to get from the prediction vector based of SATV to that based on both predictors. Again, this model is fairly intuitive and one should be able to easily see that it is equivalent to the first two models.

Table 6: Model 3: $Y = \beta_0 + \beta_1 E\_SATM + \beta_2 SATV + \varepsilon$ $R^2 = 0.1178$

	Stand_Est	Df_adj	SSM	T	df	P	Unique R^2
E_SATM	0.25511	265	10020962	4.421	265	0.000	0.065
SATV	0.22951	265	8110321	3.978	265	0.000	0.053

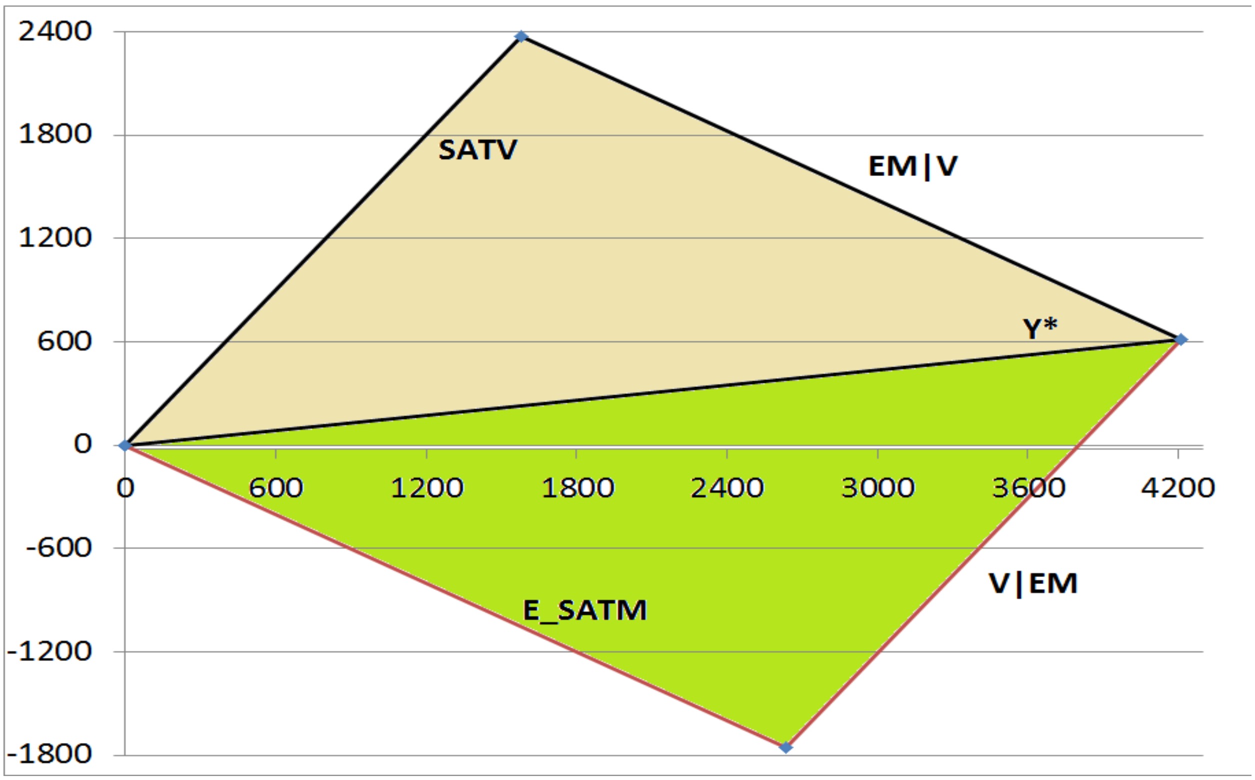

Figure 3: Geometrical Representation of Model 3

Model 4 (see Table 7) is the beginning of our less intuitive models and is not as easily seen as equivalent to the other three (Models 1-3). The predictors here are E_SATM and E_SATV. The relationship of SATV is partialled from the first, E_SATM and the relationship of SATM is partialled from the second, E_SATV. Whatever the predictors share in common seems to be lost from Model 4. Given that E_SATM is orthogonal to SATV, it then is whatever is spanned by the space that is not SATV. Thus, it is analogous in prediction power to SATM. Likewise, E_SATV is whatever is in the space unrelated to SATM which is SATV. Note that E_SATV and E_SATM create a suppressor effect (sum of unique contributions above and beyond the other exceed the sum of their sole contributions). Our results for this model tell us that everything in the spaced spanned by our predictors that is not SATM and everything that is not SATV are both significant. Given the significance of SATM, we already know that there is a lot left over after using SATV that is significant. Moreover, from Model 3 we found that what is in SATV that is not in SATM is also significant.

Figure 4 shows that we are still in the same space and that Y* is still in the same place in that space. We can also see that E_SATV is orthogonal to SATM (see Figures 1 and 4) and the E_SATM is orthogonal to SATV (see Figures 1 and 4). Figure 4 also allows us to see a pictorial representation of a suppressor effect. What it takes to get from E_SATV to Y*, EM|EV (the incremental length) is larger than the length of E_SATM initially. Likewise, what it takes to get from E_SATM to Y*, EV|EM is larger than E_SATV initially.

Table 7: Model 4: $Y = \beta_0 + \beta_1 E\_Math + \beta_2 E\_SATV + \varepsilon$ $R^2 = 0.1178$

	Stand_Est	Df_adj	SSM	T	df	P	Unique R^2
E_SATM	0.40827	265	17757089	5.886	265	0.000	0.115
E_SATV	0.27592	265	8110321	3.978	265	0.000	0.053

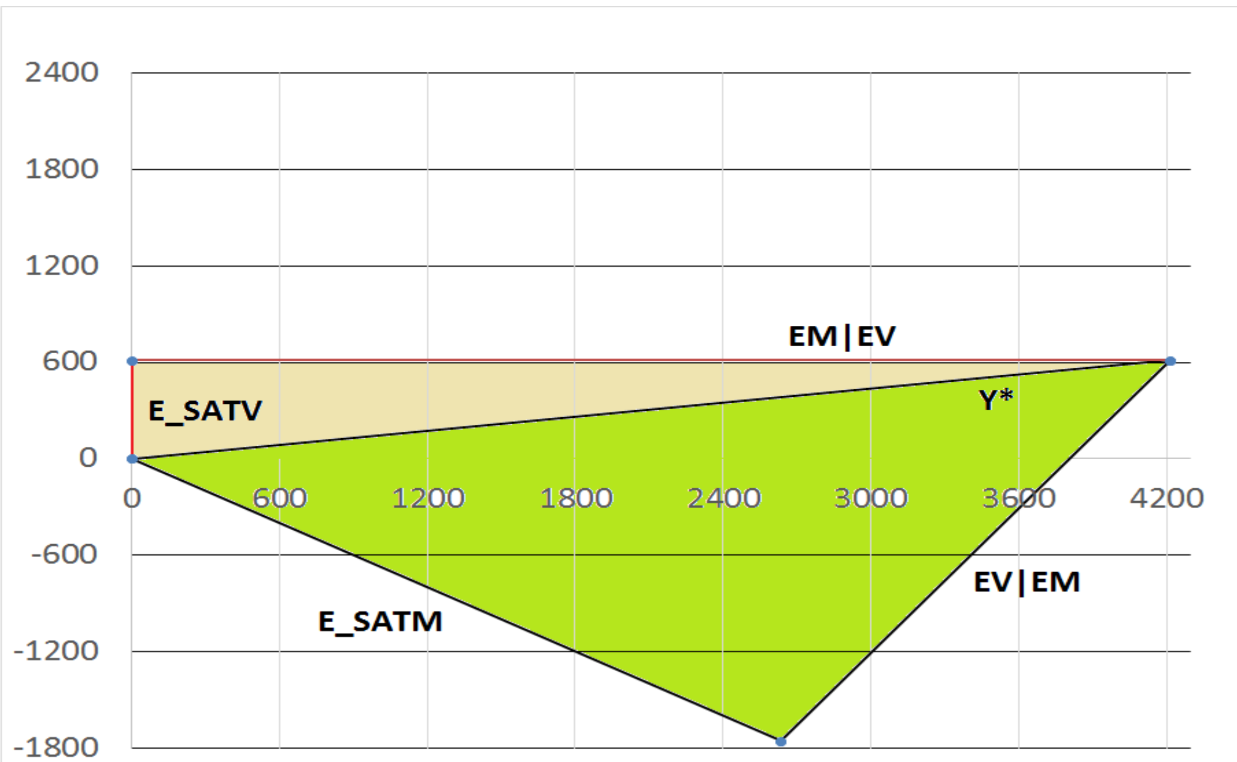

Figure 4: Geometrical Representation of Model 4

Table 8 shows the results of Model 5. While Model 4 began our presentation of less intuitive models, Model 5 takes lack of intuitiveness to a new level. The predictors in Model 5 are SATM and the error of SATM remaining after the relationship of SATV has been partialled out. Thus, E_SATM is orthogonal to SATV. On the face of it there appears to be no influence of SATV left in this model. The model consists of SATM and whatever is not SATV. Once gain SATM is a significant predictor and once SATM is in the model everything remaining in the space that is not SATV is not significant.

Figure 5 shows the corresponding geometric representation of Model 5. We have seen each of the main vectors before, SATM, E_SATM, and Y* and they are still in their same places. A basic law of geometry now becomes more apparent. Any two vectors in a 2-dimensional plane will define that plane. We have already shown that all of these vectors are coplanar (SATV, SATM, E_SATM, E_SATV, and Y*). Thus, any two of the four vectors can be used to predict Y* equally as well. The only difference will be as we have seen above; the configuration of the length of the vectors of the predictors and the additional lengths necessary to end at Y*.

Table 8: Model 5: $Y = \beta_0 + \beta_1 SATM + \beta_2 E\_SATM + \varepsilon$ $R^2 = 0.1178$

	Stand_Est	Df_adj	SSM	T	df	P	Unique R^2
SATM	0.41347	265	8110321	3.978	265	0.000	0.053
E_SATM	-0.08881	265	374193	-0.854	265	0.394	0.002

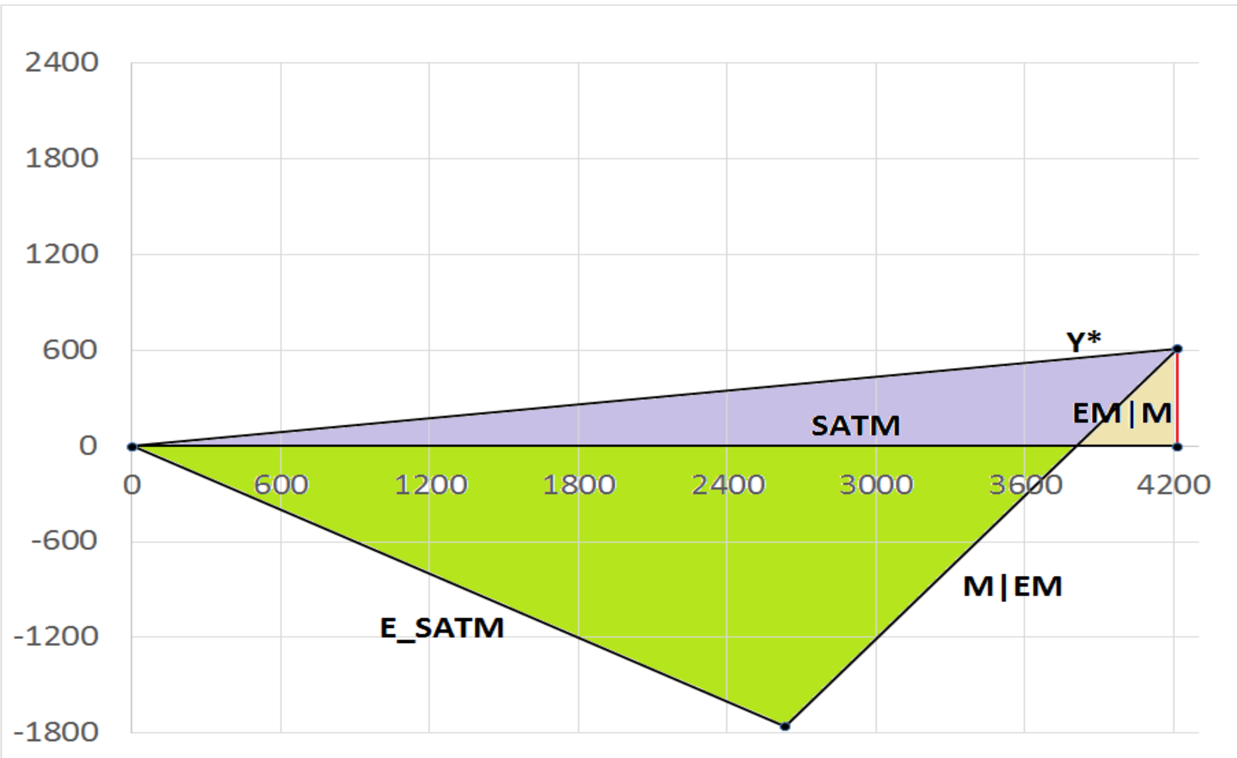

Figure 5: Geometrical Representation of Model 5

Results for the final model are shown in Table 9. This model has SATV and everything that is not SATM as predictors. Given the relationship between SATV and GPA for STEM majors, there appears to be a significant amount of variance that is not SATV and not SATM that is still useful in predicting GPA. Hence, E_SATV is significant. This final model is also a suppressor model where the additional sums of squares (Type 3) are larger than the original sums of squares.

Table 9: Model 6: $Y = \beta_0 + \beta_1 SATV_2 + \beta_2 E\_SATV + \varepsilon$ $R^2 = 0.1178$

	Stand_Est	Df_adj	SSM	T	df	P	Unique R^2
SATV	0.6118	265	17757089	5.886	265	0.000	0.115
E_SATV	-0.4596	265	10020962	-4.421	265	0.000	0.065

Figure 6 shows the geometric representation of Model 6. We have seen all of the predictors previously and note that they are coplanar and that all of the vectors are still in their same positions. Given that this is a suppressor model we also note that V|EV is greater than SATV and EV|V is greater than E_SATV. Finally, Figure 7 shows all of the vectors and their relationship to Y*. The last plot clearly shows the vectors to be coplanar and gives insight on what it means to span a plane. Implications of this insight will be discussed in the conclusions.

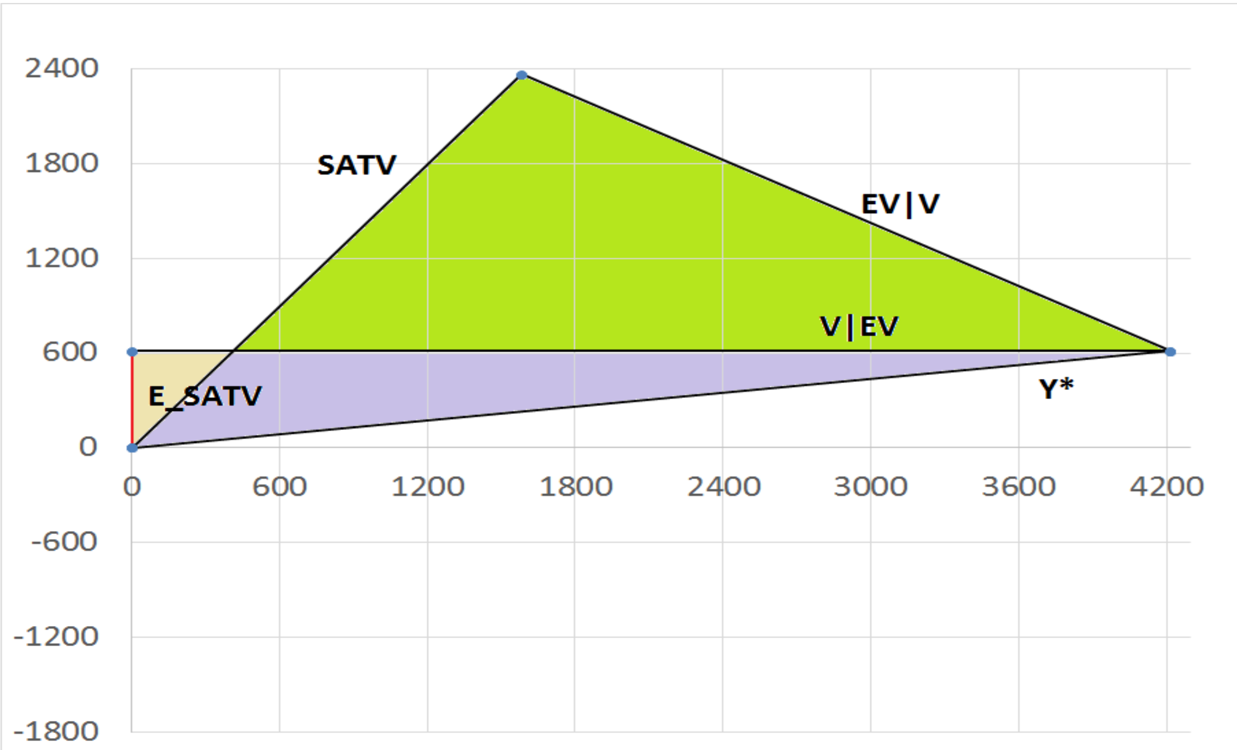

Figure 6: Geometrical Representation of Model 6

What do we know from our analyses? 1) SATV and SATM are correlated. 2) Both SATV and SATM are significantly related to the criterion, GPA for students in STEM majors. 3) SATM is more predictive of GPA for STEM majors than SATV. 4) While the incremental part of SATM (above and beyond SATV) is still a significant predictor of GPA for STEM majors, the incremental part of SATV above SATM is not. 5) There appears to be a significant verbal component that helps to predict GPA for STEM majors which is why the initial correlation of SATV and GPA is significant. However, this verbal component is also related to SATM which is the reason that SATV is no longer significantly related to GPA when SATM is also in the model. The final conclusion is more apparent as a result of the Total Information Regression Analysis.

4. Methodological Conclusion and Extension

We began this article referring to the regression approach as Total Information Regression Analysis. Note that the space spanned by the hyperplane of the predictors contains the total amount of information in the predictors and all combinations of residual vectors, solely and simultaneously. We have given a simple example using two predictors and their errors after predicting each variable with the other. Thus, we could interpret our predictors in terms of the original variables as well as in terms of residual vectors where each variable is cleansed of the relationship of the other. These predictors have different meanings and thus address different aspects of a theory. In generalizing the results of the figures one can also see that there are an infinite number of vectors that can be used as predictors (any two vectors in the plane). All of these vectors will be in the space of the original predictors. The choice of a particular vector, however, would have immediate implications relative to theory. The vectors we used were the four easiest given our two-predictor original model (original variables and residual variables after predicting one of the variables from the other).

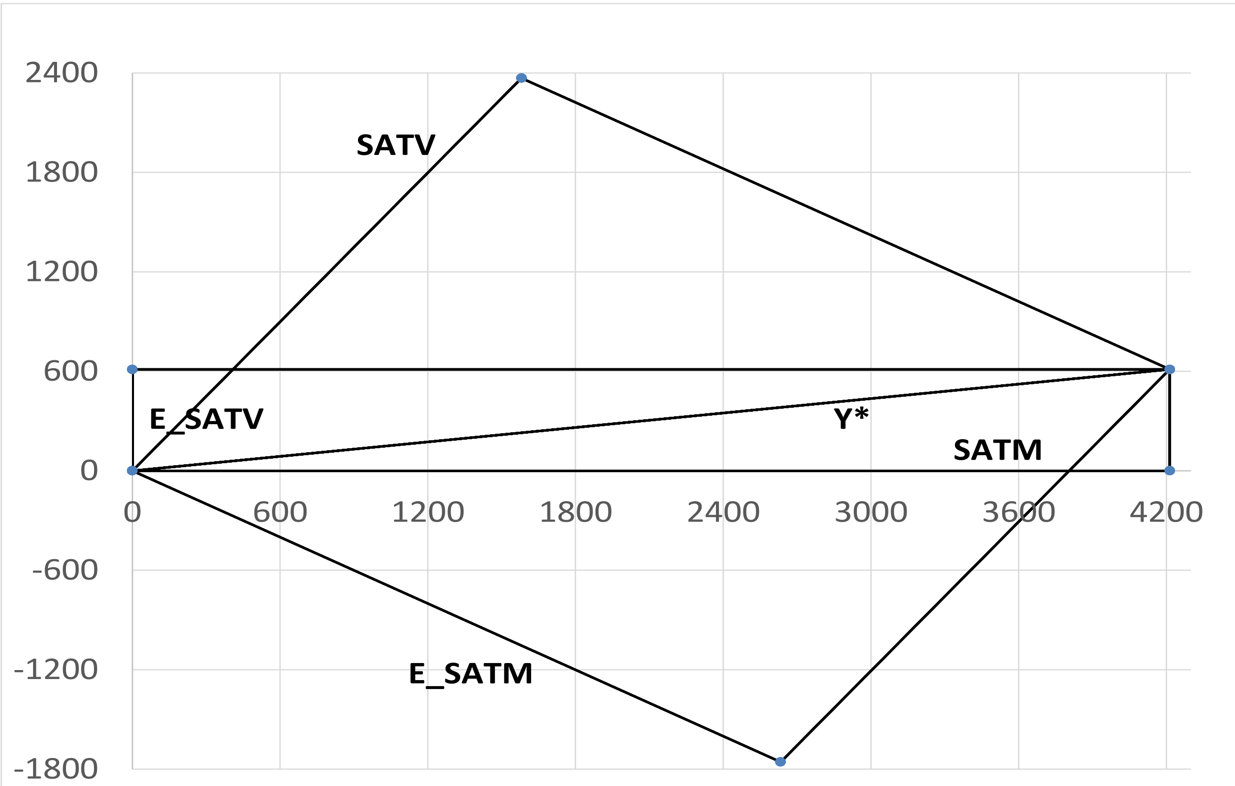

Figure 7: Geometrical Representation of all Predictors, SATV, SATM, E_SATV, and E_SATM with the Predicted Value for the 2-Predictor Model, Y*

Generalizing the approach to more predictors would lead to a host of more predictors that are available. We can make a multitude of residual vectors where we partial from 1 to k-1 (for k predictors) of the other predictors from each predictor. For example with 3 predictors we would have each of the original predictors (3), 6 error vectors were we partial one of the other two variables from a target variable in turn, and 3 residual vectors where we partial both of the other variables from a target variable For a 2-predictor model we had 4 predictors as shown above. The number has grown to 12 different predictors for a 3 predictor model. For this 3-predictor model we can use any 3 of the 12 vectors to define an equivalent model. This will lead to the possibility of 220 different models, each testing different aspects of the original variables and residual vectors. Thus, while the overall models are equivalent, each set of predictors will lead to different interpretations of the predictors and their significance. While the original 3 variables span the entire space of the 12, one will need the separate vectors to provide for different interpretations of which pieces of the original and residual variables are significant solo, in combination, and/or with different variables partialled out.

The implications of what it means to span a space are apparent from the figures. While most investigators make use of only the vectors defined by the original variables, we now see that an infinite number of vectors are available for our use and that there is more information available in the space of the predictors than typically used. We can easily define a new set of variables contained in the same space via partialling out the effects of any combination of the original predictors from each of the original predictors in turn creating a host of residual vectors. Note that this will yield new variables that allow us to address different aspects of our theory as we can test the effects of having a given predictor with the effects of any and all of the variables partialled from it. Simply using the original variables will allow us to do a regression. The ability to use the original variables in conjunction with a host of residual vectors will allow us to perform a Total Information Regression analysis. From using merely the original variables that allow us information from only k vectors in our space, one can now make use of much more information. While there are an infinite number of vectors available to us, the original and residual variables we derive provide a finite set that allows us to interpret our results.

References

Hakstian, R. A. & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology, 70*, 657-669.

Maxwell, S. E. and Delaney, H. D. (1990). *Designing experiments and analyzing data*. Delmont, CA: Wadsworth.

Schey, H. M. (1993). The relationship between the magnitude of SSR(X2) and SSR(X2|X1): A geometric description. *The American Statistician, 47*, 26-30.