

Estimation of Curvilinear Effects in SEM

Supplement to *Principles and Practice of Structural Equation Modeling* (3rd ed.)

Rex B. Kline, September 2009

Curvilinear Effects of Observed Variables

Consider the data set in Table 1. The scatterplot for the variables *X* and *Y* is presented in Figure 1, and it clearly shows that the relation between these variables is curvilinear. Specifically, it is quadratic because scores on *Y* decline and then rise as scores on *X* increase. Regressing *Y* on *X* yields a near-zero Pearson correlation of $-.048$ and the unstandardized regression equation presented next and also illustrated in Figure 1:

$$\hat{Y} = -.023 X + 10.495$$

This regression equation has poor fit because it reflects only the linear aspect of the relation between *X* and *Y*, which is slight for these data. To also represent the quadratic trend, all that is needed is to create the power term X^2 and then regress *Y* on both *X* and X^2 . The presence of X^2 in the equation adds one bend to the regression line, and its unstandardized regression coefficient indicates the degree of the quadratic aspect of *X*'s relation to *Y*, controlling for the linear effect. The multiple regression coefficient with both *X* and X^2 in the equation for the data in Table 1 is $.927$, and the unstandardized equation for the quadratic regression curve presented next and also depicted in Figure 1 clearly fits the data better than the linear prediction equation:

$$\hat{Y} = .083 X^2 - 2.714 X + 29.486$$

Insert Table 1 about here

Even higher-order curvilinear relations can be represented with the appropriate power

term. For example, the term X^3 represents the cubic relation of X to Y (assuming that X and X^2 are also in the equation), and so on. However, it is rarely necessary to estimate curvilinear relations beyond a quadratic one in behavioral data. Note that standardized regression coefficients (beta weights) do not have the standard interpretation for power terms. Some authors, such as J. Cohen, J., P. Cohen, West, and Aiken (2003), suggest that researcher should consider instead whether adding the power term to the equation results in an appreciable increase in the overall R^2 .

 Insert Figure 1 about here

Centering

A problem that can occur when analyzing power terms that represent curvilinear effects is extreme collinearity. This is because correlations between product terms and their constituent variables can be so high that the analysis can fail or the results are unstable. For example, the Pearson correlation between the power term X^2 and the original variable X is .986 for the data set in Table 1. One way to address this problem is to center the original variables before calculating product terms based on them. For example, the correlation between variables X and X^2 after centering the former before creating the latter for the data in Table 1 is practically zero ($r = .016$), much lower than for the original (uncentered) scores (.986).

Representation in Path Models

Curvilinear effects of observed variables are represented in path models with the appropriate product terms and all constituent variables. Consider the path model in Figure 2. This model depicts X and the power term X^2 as exogenous variables and Y as endogenous. The unanalyzed

association between the two exogenous variables in the figure accounts for any correlation between variable X and the power term X^2 . The unstandardized path coefficient for the path $X \rightarrow Y$ estimates the linear effect of X , and the unstandardized coefficient for the path $X^2 \rightarrow Y$ estimates the quadratic effect of X , each controlling for the other effect.

 Insert Figure 2 about here

Curvilinear Effects of Latent Variables

In the indicant product approach in SEM, power terms are specified as multiple indicators of latent variables that represent curvilinear effects. Consider the standard structural-regression (SR) model in Figure 3(a). The path $A \rightarrow Y$ represents the linear effect of latent variable A on the observed endogenous variable Y . (The rationale outlined here also applies to the use of multiple indicators to measure an endogenous construct.) The measurement model for the indicators of factor A can be represented with the following structural equations:

$$X_1 = A + E_1 \tag{1}$$

$$X_2 = \lambda_2 A + E_2$$

where the loading of X_1 is fixed to 1.0 scale the factor and the loading of X_2 is a free parameter represented by the term λ_2 . The only other parameters of this model are the variances of A and the error terms, E_1 and E_2 . We assume next that the latent variables just mentioned are all independent with means of zero (i.e., they are in mean-deviated form).

 Insert Figure 3 about here

Suppose that a researcher wished to estimate the quadratic effect of factor A on Y . This calls for adding to the model of Figure 3(a) the latent product variable A^2 that represents this effect, which is estimated by the coefficient for the path $A^2 \rightarrow Y$. Like its linear counterpart, A^2 is a latent variable measured indirectly only through its indicators. The indicators of A^2 are the product indicators

$$X_1^2, X_2^2, \text{ and } X_1 X_2$$

Note that the term $X_1 X_2$ does not here represent an interactive effect because its components, X_1 and X_2 , are specified to measure the same factor. By squaring or taking the product of the corresponding expressions in Equation 1, the equations of the measurement model for the product indicators are

$$X_1^2 = A^2 + 2AE_1 + E_1^2 \quad (2)$$

$$X_2^2 = \lambda_2^2 A^2 + 2\lambda_2 AE_2 + E_2^2$$

$$X_1 X_2 = \lambda_2^2 A^2 + \lambda_2 AE_1 + AE_2 + E_1^2 E_2^2$$

These equations (2) show that the product indicators actually load on a total of *six* different latent product variables, including

$$A^2, AE_1, AE_2, E_1^2, E_2^2, \text{ and } E_1 E_2$$

The latter three terms just listed are the residual terms for the product indicators. Note that all the factor loadings of the measurement model for the product indicators are either constants or functions of λ_2 , the loading of the X_2 on factor A . For example, the loading of X_1^2 on factor AE_1

equals 2.0, and the loading of X_2^2 on factor A^2 equals λ_2^2 . Thus, no new factor loadings need to be estimated for the product indicators.

The other parameters of the measurement model for the product indicators are the variances and covariances of the six latent product variables implied by Equation 2. Assuming normal distributions for the nonproduct latent variables A , E_1 , and E_2 , it can be shown that the six latent product terms in Equation 2 and factor A are all mutually uncorrelated (i.e., their covariances are zero; see Kenny & Judd, 1984, pp. 210). Under the same assumption, it also true that the variances of the latent product terms in Equation 2 can be expressed as functions of the variances of the nonproduct latent variables A , E_1 , and E_2 as follows:

$$\begin{aligned} \sigma_{A^2}^2 &= 2(\sigma_A^2)^2 & \sigma_{E_1 E_2}^2 &= \sigma_{E_1}^2 \sigma_{E_2}^2 \\ \sigma_{E_1^2}^2 &= 2(\sigma_{E_1}^2)^2 & \sigma_{AE_1}^2 &= \sigma_A^2 \sigma_{E_1}^2 \\ \sigma_{E_2^2}^2 &= 2(\sigma_{E_2}^2)^2 & \sigma_{AE_2}^2 &= \sigma_A^2 \sigma_{E_2}^2 \end{aligned} \quad (3)$$

For example, the variance of the latent product term A^2 equals two times the squared variance of factor A , and the variance of the term $E_1 E_2$ equals the product of the measurement error variances for the nonproduct indicators X_1 and X_2 . Thus, no new variances need to be estimated. This means that the parameters of the measurement model for the product indicators implied by Equations 2 and 3 are theoretically identified, given all of the assumptions stated to this point. Presented in Figure 3(b) is the whole SR model that includes the measurement model for the nonproduct indicators of A , the measurement model for the product indicators of A^2 , and the structural model for predicting Y . The coefficient for the path $A^2 \rightarrow Y$ estimates the latent quadratic effect.

Estimation with the Kenny-Judd Method

Application of the Kenny-Judd method (Kenny & Judd, 1984) to estimate a latent quadratic effect is demonstrated next. I fitted with Mplus 5.2 the model of Figure 3(b) to the covariance matrix generated by Kenny and Judd (1984) for a hypothetical sample of 500 cases that is presented in Table 2. Because Kenny and Judd (1984) used a generalized least squares (GLS) estimator in their original analysis of these data, I specified the same estimator in this analysis with Mplus. Presented in Appendix A is Mplus syntax that specifies the model, data, and all constraints of the Kenny-Judd method for this example. In order to apply nonlinear constraints in Mplus, it is generally necessary to label in syntax the corresponding parameters. Next, the relevant constraints are listed under the “Model Constraints” heading of the command file. With a total of 6 observed variables (Y , X_1 , X_2 , and three product indicators), there are a total of $6(7)/2$, or 21 observations available for the analysis. There are a total of 7 free parameters, including

1. 3 variances (of A , E_1 , and E_2) and 1 factor loading ($A \rightarrow X_2$) for the measurement model of the nonproduct indicators; and
2. 2 directs on Y ($A \rightarrow Y$, $A^2 \rightarrow Y$) and 1 disturbance variance of Y for the structural model.

There are no free parameters for the measurement model of the product indicators, so $df_M = 14$.

 Insert Table 2 about here

Estimation in Mplus converged to an admissible solution. Values of selected fit statistics presented next indicate satisfactory overall fit of the model to the data; the 90% confidence interval for the RMSEA is reported in parentheses:

$$\chi_M^2(14) = 18.418, p = .188$$

$$\text{RMSEA} = .025 (0-.053), p_{\text{close-fit } H_0} = .924$$

$$CFI = .994; SRMR = .039$$

Reported in Table 3 are the GLS estimates of the free parameters for the model in Figure 3(b). These results are very similar to those reported by Kenny and Judd (1984) in their original analysis. The linear and quadratic effects of factor A together explain 73.4% of the total variance in Y . The form of the functional relation between Y and A is

$$\hat{Y} = .247 A - .500 A^2 \quad (4)$$

where the weights are the unstandardized path coefficients. The equation just presented describes an inverted parabola concerning the relation between A and Y . You can create a visual plot of this relation using a general parabola equation plotter that is freely available over the Internet.¹ A screenshot of the parabola generated by the plotter on the site listed in the footnote for Equation 4 is presented in Figure 4.

 Insert Table 3 about here

See Marsh, Wen, and Hau (2006) for descriptions of alternative methods to estimate curvilinear effects of latent variables. These same methods can also be applied to the analysis of interactive effects of latent variables. A computationally simpler version of Klein and Moosbrugger's (2000) latent moderated structural equations (LMS) method, which does not require nonlinear constraints or the creation of power terms, is incorporated in Mplus 5.2 with special compact syntax for analyzing models with latent curvilinear or interactive effects (Klein and Muthén, 2007).

¹sasked.gov.sk.ca/docs/math30/conics/conics/laboratory/parABC.html

Insert Figure 4 about here

References

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.
- Klein, A., & Moosbrugger, A. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.
- Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*, 647–673.
- Marsh, H. W., Wen, Z., & Hau, K.-T. (2006). Structural equation modeling of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 225-265). Greenwich, CT: IAP.

Appendix A

Mplus Syntax for Estimation of the Model in Figure 3(b) with the Data in Table 2

Input file

TITLE: Kenny-Judd for latent quadratic effect

DATA:

FILE IS "quadratic-kenny-judd-mplus.dat"; TYPE IS COVARIANCE;

NGROUPS = 1; NOBSEVATIONS = 500;

VARIABLE:

NAMES ARE x1 x2 x12 x22 x1x2 y;

ANALYSIS:

TYPE IS GENERAL; ESTIMATOR = GLS;

MODEL:

! measurement model

! parameters for nonproduct indicators

! V is for variance, L for loading

A (V_A);

x1 (V_E1);

x2 (V_E2);

A BY x1

 x2 (L_x2);

! parameters for product indicators

A2 (V_A2);

x12 (V_Ex12);

x22 (V_Ex22);

x1x2 (V_Ex1x2);

```

A2 BY x12 ! fixed to 1 by default
      x22 (L_x22_A2)
      x1x2 (Lx1x2A2);

AE1 (V_AE1);

AE1 BY x12@2 ! fixed to 2
      x1x2 (Lx1x2AE1);

AE2 (V_AE2);

AE2 BY x1x2 ! fixed to 1 by default
      x22 (Lx22AE2);

! fix all factor covariances to zero
A with A2@0; A with AE1@0;
A with AE2@0; A2 with AE1@0;
A2 with AE2@0; AE1 with AE2@0;

! structural model
y ON A A2;

MODEL CONSTRAINT:

! variance constraints for product factors
V_A2 = 2 * V_A**2;
V_Ex12 = 2 * V_E1**2;
V_Ex22 = 2 * V_E2**2;
V_Ex1x2 = V_E1 * V_E2;
V_AE1 = V_A * V_E1;
V_AE2 = V_A * V_E2;

! loading constraints for product indicators
! not already fixed to equal a constant

```

$L_x22_A2 = L_x2^{**2};$

$Lx1x2A2 = L_x2;$

$Lx1x2AE1 = L_x2;$

$Lx22AE2 = 2 * L_x2;$

OUTPUT: SAMPSTAT RESIDUAL STDYX;

Contents of file "quadratic-kenny-judd-mplus.dat"

1.150

.617 .981

-.068 -.025 2.708

.075 .159 .729 1.717

.063 .065 1.459 1.142 1.484

.256 .166 -1.017 -.340 -.610 .763

TABLE 1. Data Set for a Quadratic Effect of Observed Variables

Case	Predictors		Criterion
	X	X^2	Y
A	7	49	14
B	10	100	11
C	13	169	9
D	15	225	9
E	17	289	5
F	19	361	8
G	23	529	11
H	25	625	14

TABLE 2. Input Data (Covariances) for Analysis of a Model with a Quadratic Effect of a Latent Variable

Variable	1	2	3	4	5	6
1. X_1	1.150					
2. X_2	.617	.981				
3. X_1^2	-.068	-.025	2.708			
4. X_2^2	.075	.159	.729	1.717		
5. $X_1 X_2$.063	.065	1.459	1.142	1.484	
6. Y	.256	.166	-1.017	-.340	.610	.763

Note. These data for a hypothetical sample are from Kenny and Judd (1984, p. 204); $N = 500$.

TABLE 3. Generalized Least Squares Estimates for a Model with a Quadratic Effect of a Latent Variable

Parameter	Unstandardized	SE	Standardized
<u>Loadings</u>			
$A \rightarrow X_1$	1.000 ^a	—	.928
$A \rightarrow X_2$.624	.025	.645
<u>Variances</u>			
A	.989	.045	1.000
E_1	.160	.033	.860
E_2	.540	.019	.417
D_Y	.199	.036	.266
<u>Direct effects</u>			
$A \rightarrow Y$.247	.028	.284
$A^2 \rightarrow Y$	-.500	.037	-.808

Note. Standardized estimates for error terms are proportions of unexplained variance.

^aNot tested for statistical significance. For all other unstandardized estimates, $p < .01$.

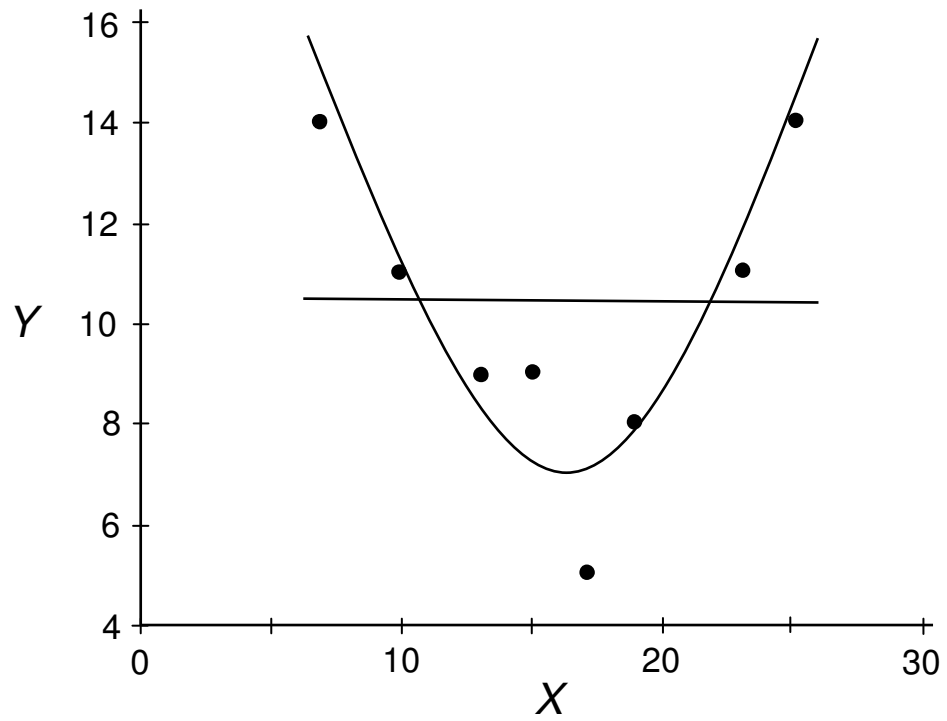


FIGURE 1. Scatterplot for data set (a) in Table 12.1 with the linear and curvilinear (quadratic) regression lines.

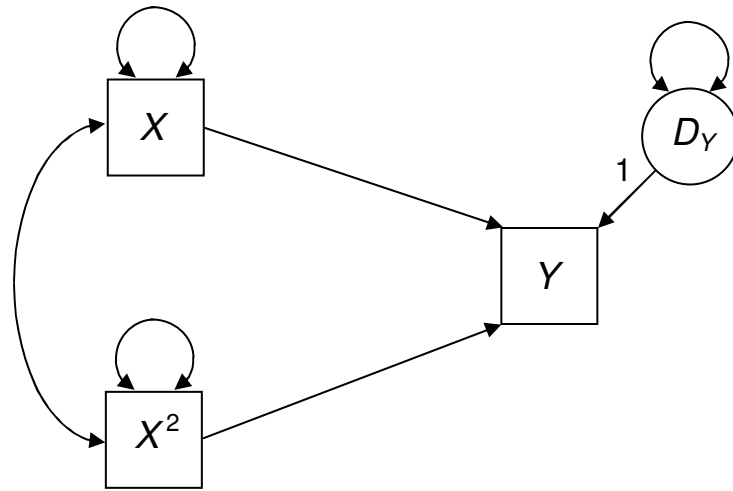


FIGURE 2. Path model with terms that represents the linear and quadratic effects of variable X .

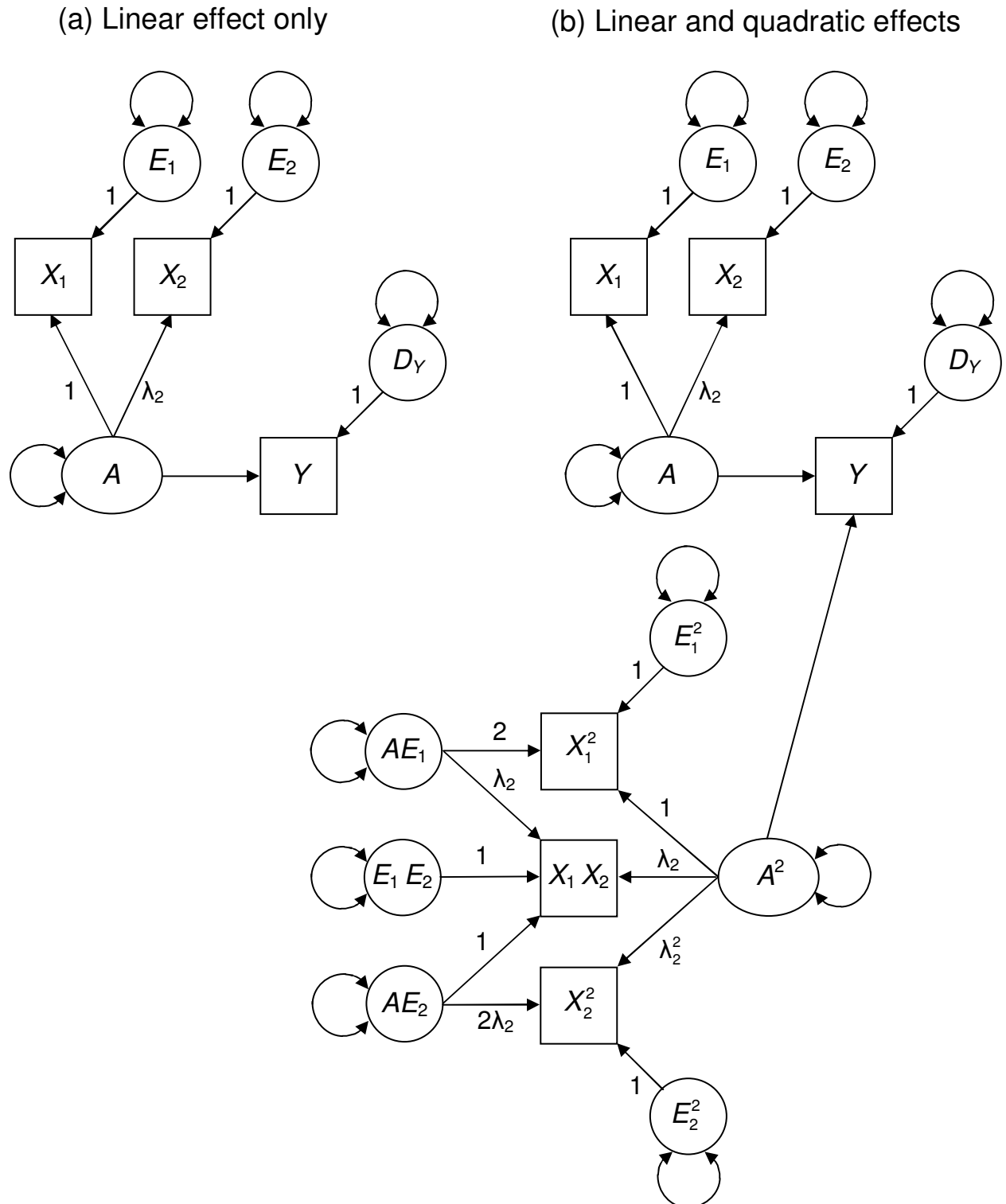


FIGURE 3. (a) A model with a linear effect of latent variable A only. (b) The corresponding model with linear and quadratic effects of latent variable A .

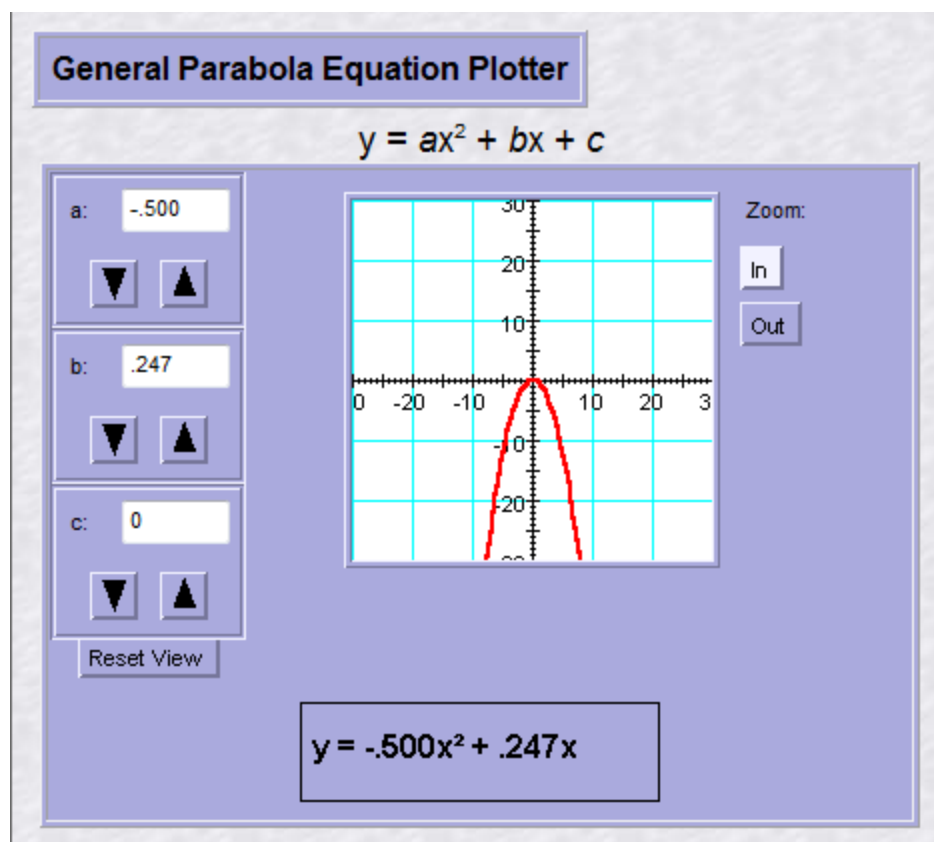


FIGURE 4. Plot of Equation 4 for the model in Figure 3(b) estimated with the data in Table 2.