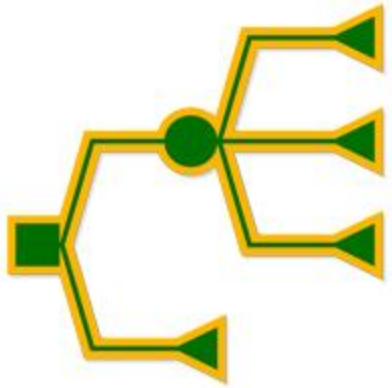
Better Exponential Curve Fitting Using Excel

Mike Middleton
DSI 2010 San Diego

Michael R. Middleton, Ph.D.
Decision Toolworks
Mike@DecisionToolworks.com
415.310.7190



Background

- The exponential function, Y=c*EXP(b*x), is useful for fitting some non-linear single-bulge data patterns.

- In Excel, you can create an XY (Scatter) chart and add a best-fit "trendline" based on the exponential function.

- **Problem**: Regarding the fitted curve for Excel's Exponential Trendline,

  (1) the reported value for R Squared is incorrect, and

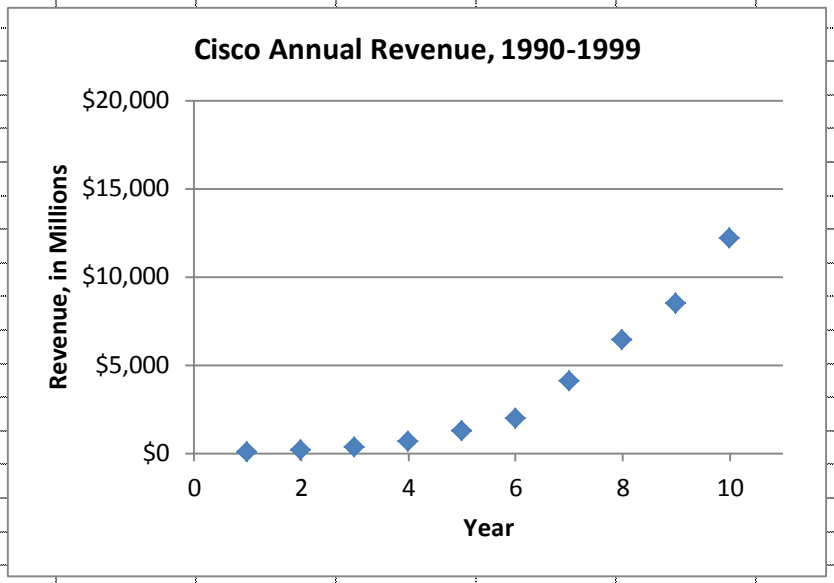
  (2) the fitted values do not minimize Sum of Squared Deviations.

Cisco Revenue Example

- Data from example originally presented in Winston (2004)
- Model for growth of Cisco revenue during 1900-1999
- Potentially useful for projecting revenues and determining company value
- For 1900-1999, Cisco revenue seems to grow by approximately the same percentage each year
- The exponential function, Y=c*EXP(b*X), has the property that for each unit increase in X the value of Y increases by a constant percentage

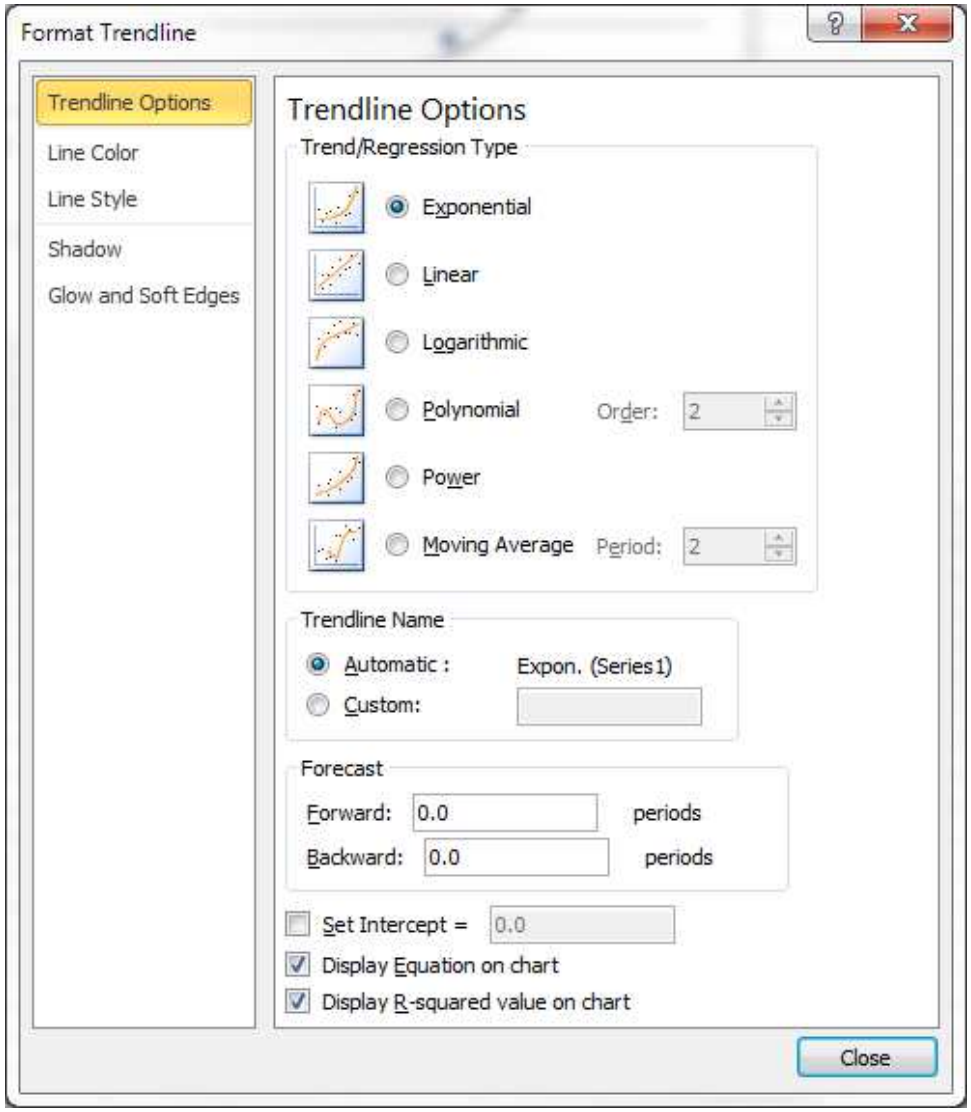
Cisco Data and XY Chart

	A	B	C	D	E	F	G	H	I
1		$ Millions							
2	Year	Revenue							
3	X	Y							
4	1	$70							
5	2	$183							
6	3	$340							
7	4	$649							
8	5	$1,243							
9	6	$1,979							
10	7	$4,096							
11	8	$6,440							
12	9	$8,459							
13	10	$12,154							
14									
15									
16									
17									
18									



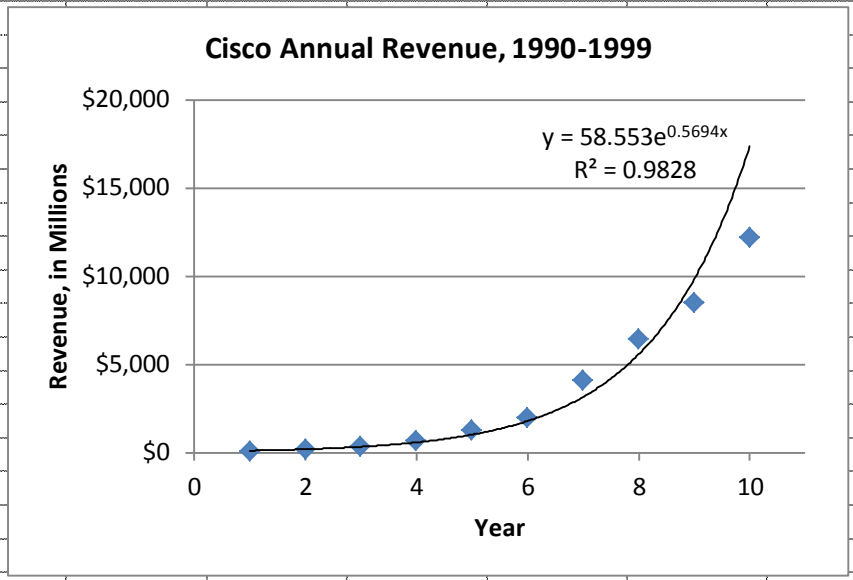
- In Excel 2010, select data A4:B13. Insert XY Scatter chart. Use Chart Tools Layout to add chart title and axes titles.
- Right-click a data point to select the data series, and choose Add Trendline from the shortcut menu.

Trendline Dialog Box



Excel Chart with Exponential Trendline

	A	B
1		$ Millions
2	Year	Revenue
3	X	Y
4	1	$70
5	2	$183
6	3	$340
7	4	$649
8	5	$1,243
9	6	$1,979
10	7	$4,096
11	8	$6,440
12	9	$8,459
13	10	$12,154
14		
15		
16		
17		
18		



Next, compute the fitted values for Y, and use worksheet functions and formulas to compute the actual value of R Squared

Actual R Squared for Exponential Trendline

	A	B	C	D	E	F	G	H
1		$ Millions			Exponential Trendline			
2	Year	Revenue						
3	X	Y	Fitted Y					
4	1	$70	$103		SS Total	156,733,316		Total SS
5	2	$183	$183		SS Regression	125,667,007		Explained SS
6	3	$340	$323		SS Residual	31,066,309		Unexplained SS, SSD
7	4	$649	$571					
8	5	$1,243	$1,009		R Squared	0.802		Explained SS / Total SS
9	6	$1,979	$1,783					
10	7	$4,096	$3,151		StDev(Residuals)	$1,763		
11	8	$6,440	$5,568					
12	9	$8,459	$9,840					
13	10	$12,154	$17,389					

Excel's Trendline reports R Squared = 0.9828

Actual R Squared = 0.802

"Approximately 80% of the variation in Y is explained by X using the fitted exponential function"

“Shortcut” Excel functions for R Squared calculations

	A	B	C	D	E	F	G	H
1		$ Millions			Exponential Trendline			
2	Year	Revenue						
3	X	Y	Fitted Y					
4	1	$70	$103		SS Total	=COUNT(B4:B13)*VARP(B4:B13)		Total SS
5	2	$183	$183		SS Regression	=F4-F6		Explained SS
6	3	$340	$323		SS Residual	=SUMXMY2(B4:B13,C4:C13)		Unexplained SS, SSD
7	4	$649	$571					
8	5	$1,243	$1,009		R Squared	=F5/F4		Explained SS / Total SS
9	6	$1,979	$1,783					
10	7	$4,096	$3,151		StDev(Residuals)	=SQRT(F6/COUNT(B4:B13))		
11	8	$6,440	$5,568					
12	9	$8,459	$9,840					
13	10	$12,154	$17,389					

Note that we cannot use Excel’s worksheet functions RSQ or PEARSON^2 or CORREL^2 to compute R Squared because those functions are based on a linear fit between Y and X.

Setup display for better fit using Excel's Solver

	A	B	C	D	E	F	G	H
1		$ Millions			Coeff c	Coeff b		
2	Year	Revenue			58.55266	0.569367		
3	X	Y	c*EXP(b*X)					
4	1	$70	$103		SS Total	156,733,316		Total SS
5	2	$183	$183		SS Regression	125,666,623		Explained SS
6	3	$340	$323		SS Residual	31,066,693		Unexplained SS, SSD
7	4	$649	$571					
8	5	$1,243	$1,009		R Squared	0.802		Explained SS / Total SS
9	6	$1,979	$1,783					
10	7	$4,096	$3,151		StDev(Residuals)	$1,763		
11	8	$6,440	$5,568					
12	9	$8,459	$9,840					
13	10	$12,154	$17,389					

Tentative values for coefficients in E2:F2 (Solver "Changing Cells")

Formula for fitted value in C4 depends on coefficients and X, copied to C5:C13

Sum of Squared Deviations formula in F6 (Solver "Objective") to be minimized

Setup formulas for better fit using Excel's Solver

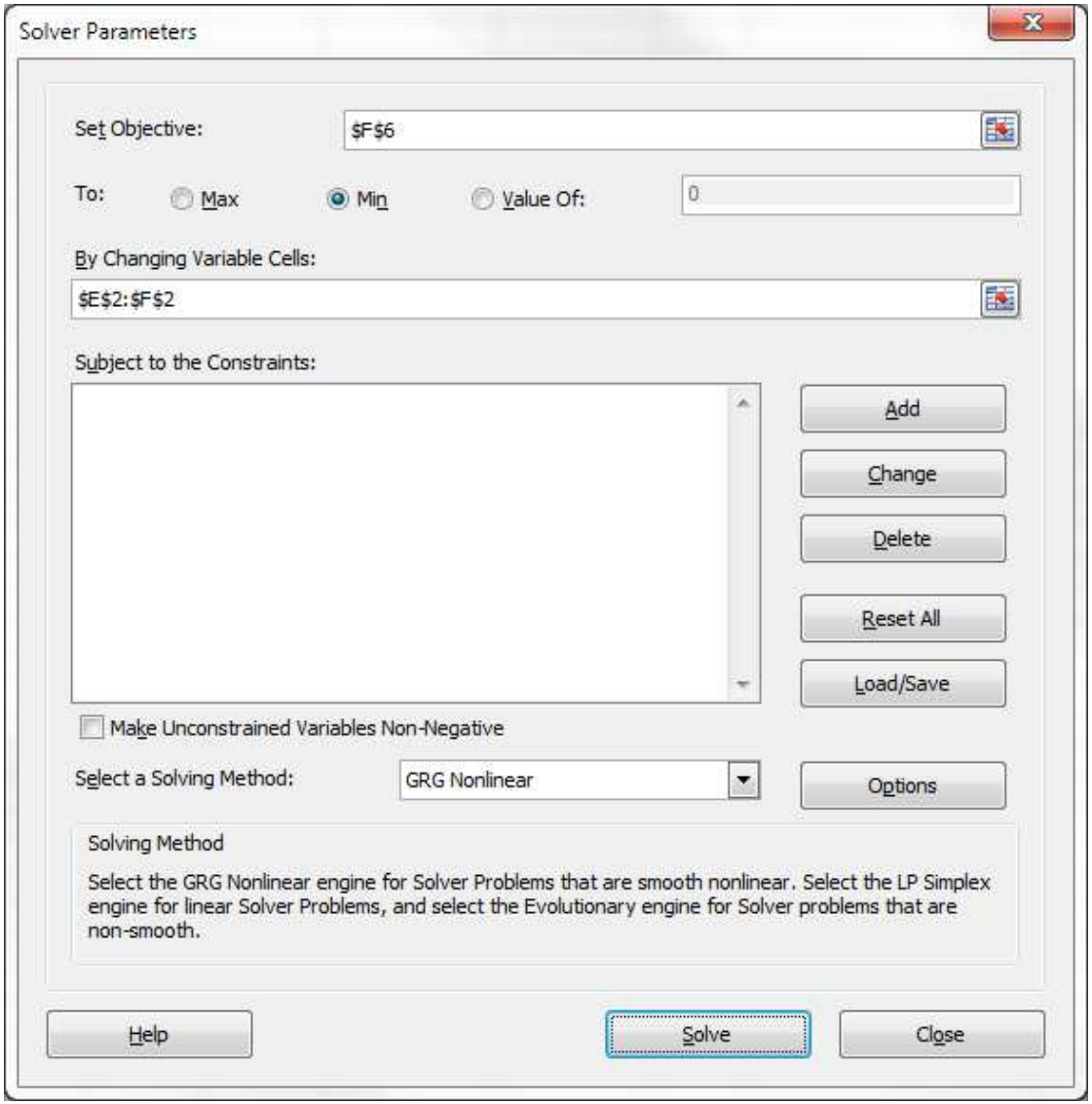
	A	B	C	D	E	F	G	H
1		$ Millions			Coeff c	Coeff b		
2	Year	Revenue			58.55266	0.569367		
3	X	Y	c*EXP(b*X)					
4	1	$70	=$E$2*EXP($F$2*A4)		SS Total	=COUNT(B4:B13)*VARP(B4:B13)		Total SS
5	2	$183	=$E$2*EXP($F$2*A5)		SS Regression	=F4-F6		Explained SS
6	3	$340	=$E$2*EXP($F$2*A6)		SS Residual	=SUMXMY2(B4:B13,C4:C13)		Unexplained SS, SSD
7	4	$649	=$E$2*EXP($F$2*A7)					
8	5	$1,243	=$E$2*EXP($F$2*A8)		R Squared	=F5/F4		Explained SS / Total SS
9	6	$1,979	=$E$2*EXP($F$2*A9)					
10	7	$4,096	=$E$2*EXP($F$2*A10)		StDev(Residuals)	=SQRT(F6/COUNT(B4:B13))		
11	8	$6,440	=$E$2*EXP($F$2*A11)					
12	9	$8,459	=$E$2*EXP($F$2*A12)					
13	10	$12,154	=$E$2*EXP($F$2*A13)					

Tentative values for coefficients in E2:F2 (Solver "Changing Cells")

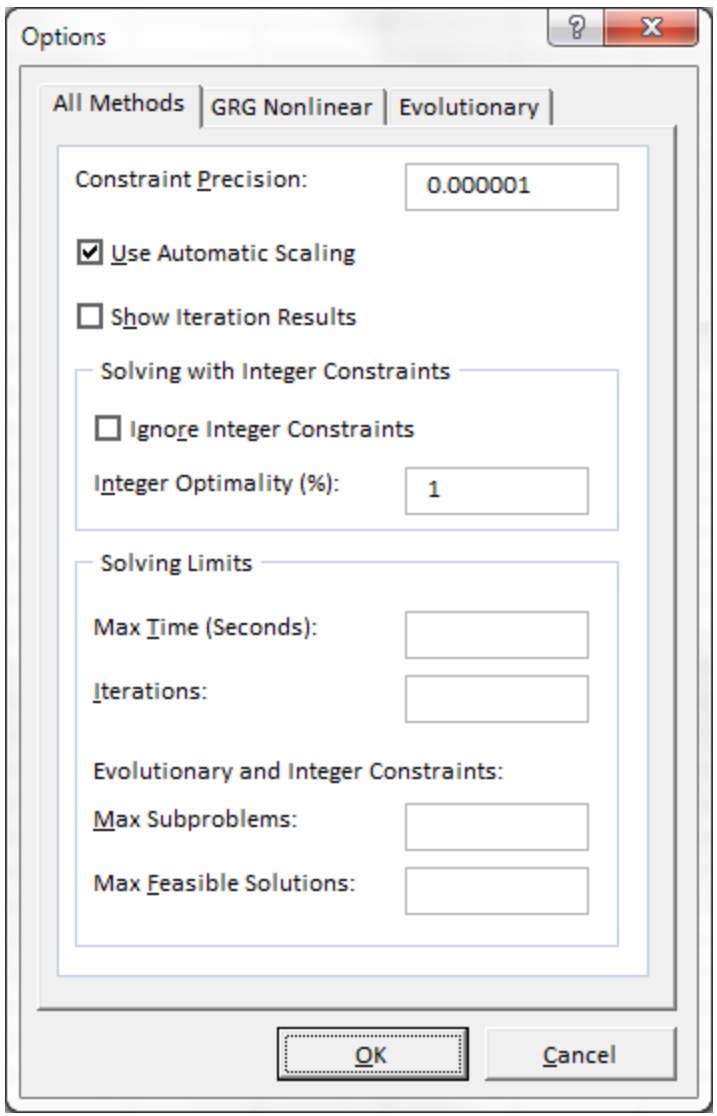
Formula for fitted value in C4 depends on coefficients and X (absolute references to E2:F2, relative reference to A4), copied to C5:C13

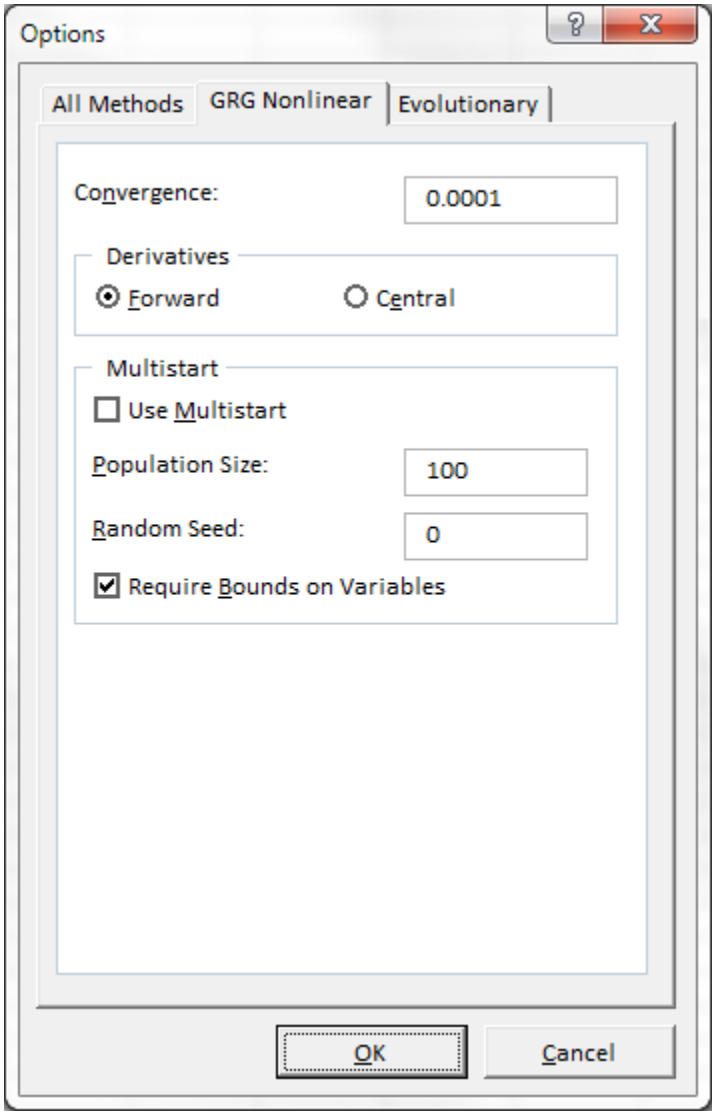
Sum of Squared Deviations formula in F6 (Solver "Objective") to be minimized

Excel 2010 Solver Parameters Dialog Box



Excel 2010 Solver Options Dialog Boxes





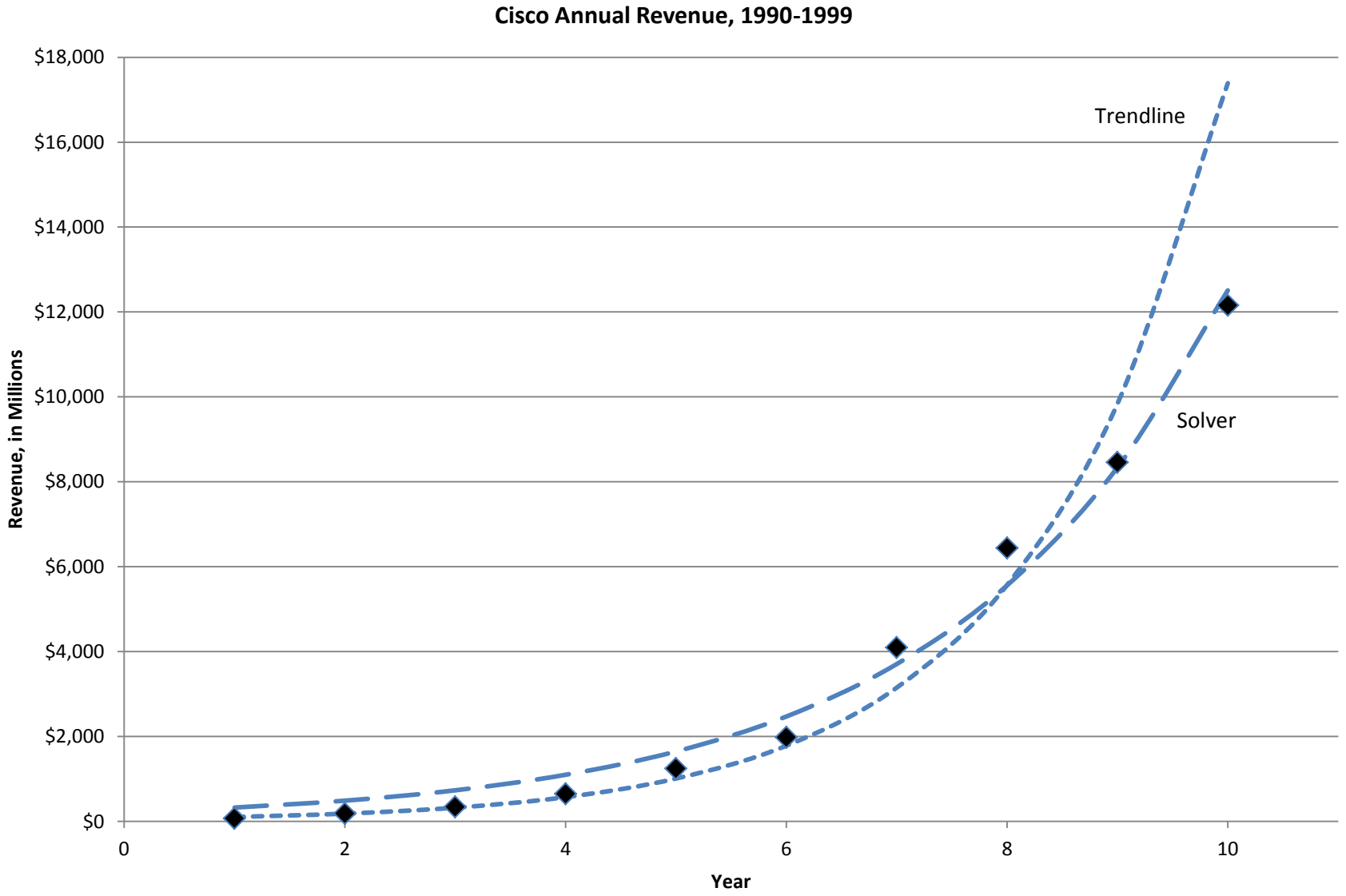
Results for Exponential Fit using Solver

	A	B	C	D	E	F	G	H
1		$ Millions			Coeff c	Coeff b		
2	Year	Revenue			217.0084285	0.40542436		
3	X	Y	c*EXP(b*X)					
4	1	$70	$325		SS Total	156,733,316		Total SS
5	2	$183	$488		SS Regression	154,746,736		Explained SS
6	3	$340	$732		SS Residual	1,986,580		Unexplained SS, SSD
7	4	$649	$1,098					
8	5	$1,243	$1,648		R Squared	0.987		Explained SS / Total SS
9	6	$1,979	$2,471					
10	7	$4,096	$3,707		StDev(Residuals)	$446		
11	8	$6,440	$5,560					
12	9	$8,459	$8,339					
13	10	$12,154	$12,509					

Excel's Trendline reported R Squared = 0.9828, but its actual R Squared = 0.802 and StDev(Residuals) = $1,763

Solver's better fit has actual R Squared = 0.987 and StDev(Residuals) = $446

Visual Comparison of Fits



Comparison of Current/Previous Ratios

	$ Millions			R^2=0.802, SD(Resid)=$1763			R^2=0.987, SD(Resid)=$446	
Year	Revenue	Actual		Trendline Exponential			Solver Fit Exponential	
X	Y	Current/Previous		Fitted Y	Current/Previous		Fitted Y	Current/Previous
1	$70			$103			$325	
2	$183	2.614		$183	1.767		$488	1.500
3	$340	1.858		$323	1.767		$732	1.500
4	$649	1.909		$571	1.767		$1,098	1.500
5	$1,243	1.915		$1,009	1.767		$1,648	1.500
6	$1,979	1.592		$1,783	1.767		$2,471	1.500
7	$4,096	2.070		$3,151	1.767		$3,707	1.500
8	$6,440	1.572		$5,568	1.767		$5,560	1.500
9	$8,459	1.314		$9,840	1.767		$8,339	1.500
10	$12,154	1.437		$17,389	1.767		$12,509	1.500
Average Ratio, 2 to 10		1.809						
Average Ratio, 3 to 10		1.708						
Average Ratio, 8 to 10		1.441						

Excel's Method for Fitting Exponential Trendline, 1 of 2

"The exponential model creates a trendline using the equation

$$y = c * e^{bx}.$$

Excel uses a log transformation of the original *y* data to determine fitted values, so the values of the dependent variable in your data set must be positive.

…

The exponential trendline feature does not find values of *b* and *c* that minimize the sum of squared deviations between actual *y* and predicted *y* (= $c * e^{bx}$). Instead, Excel's method takes the logarithm of both sides of the exponential formula, which then can be written as

$$\text{Ln}(y) = \text{Ln}(c) + b * x$$

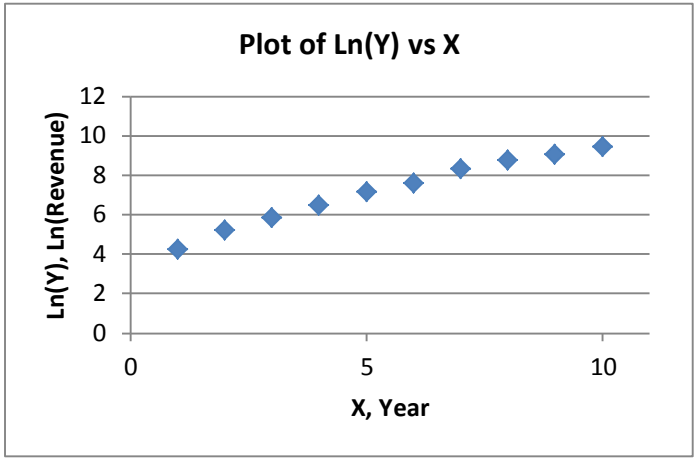
and uses standard linear regression with Ln(y) as the dependent variable and x as the explanatory variable. That is, Excel finds the intercept and slope that minimize the sum of squared deviations between actual Ln(y) and predicted Ln(y), using the formula

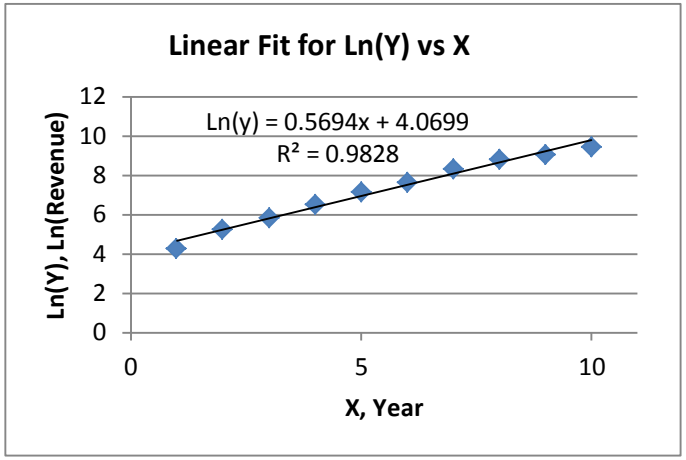
$$\text{Ln(y)} = \text{Intercept} + \text{Slope} * \text{x}.$$

Therefore, the Intercept value corresponds to Ln(*c*), and *c* in the exponential formula is equal to Exp(Intercept). The Slope value corresponds to *b* in the exponential formula." - Middleton (1995)

Excel's Method for Fitting Exponential Trendline, 2 of 2

X	Y	Ln(Y)
1	70	4.248495
2	183	5.209486
3	340	5.828946
4	649	6.475433
5	1243	7.125283
6	1979	7.590347
7	4096	8.317766
8	6440	8.770284
9	8459	9.042986
10	12154	9.405414





Y = c*EXP(b*X)

LN(Y) = LN(c) + b*X

Fit: LN(Y) = 4.0699 + 0.5694*X

b = 0.5694

c = EXP(LN(c))
c = EXP(4.0699)
c = 58.55

General Steps for Curve Fitting

Goal: explain variation in a variable of interest, Y
Prepare a histogram for Y, the dependent (or response) variable
Find data for explanatory variable(s) that make sense
Look at the data: plot XY (Scatter) charts to see relationships
Propose a functional form for the relationship, based on
    knowledge of the underlying process,
    visual examination of the plot, parsimony, etc.
Determine values for the parameters of the function
    best fit, minimize sum of squared deviations
    answers the question: What is the relationship?
Perform diagnostics, e.g., R Squared, StDev(Residuals), etc.
    answers the question: How good is the relationship?
Use the function
    prediction for cross-sectional data, mostly interpolation
    forecasts for time-series data, mostly extrapolation

Summary of Excel Trendline Options

- Exponential: Y=c*EXP(b*X), transforms data before fit, not the best fit, inaccurate R Squared
- Linear: $Y=b_0+b_1*X$, OK
- Logarithmic: Y = c*LN(X)+b, OK
- Polynomial: $Y=b_0+b_1*X_1+b_2*X_2+\ldots$, OK
- Power: Y = c*X^b, transforms data before fit, not the best fit, inaccurate R Squared
- Moving Average: OK, but non-standard diagnostics

Excel's Method for Fitting Power Trendline

The power model creates a trendline using the equation

$$y = c * x^b.$$

Excel uses a log transformation of the original *x* and *y* data to determine fitted values, so the values of both the dependent and explanatory variables in your data set must be positive.

…

The power trendline feature does not find values of *b* and *c* that minimize the sum of squared deviations between actual *y* and predicted *y* (= $c * x^b$). Instead, Excel's method takes the logarithm of both sides of the power formula, which then can be written as

$$\text{Ln}(y) = \text{Ln}(c) + b * \text{Ln}(x),$$

and uses standard linear regression with Ln(*y*) as the dependent variable and Ln(*x*) as the explanatory variable. That is, Excel finds the intercept and slope that minimize the sum of squared deviations between actual Ln(*y*) and predicted Ln(*y*), using the formula

$$\text{Ln}(y) = \text{Intercept} + \text{Slope} * \text{Ln}(x).$$

Therefore, the Intercept value corresponds to Ln(*c*), and *c* in the power formula is equal to Exp(Intercept). The Slope value corresponds to *b* in the power formula.

References

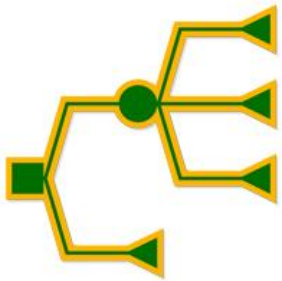
- Middleton, M.R. 1995. Data Analysis Using Microsoft Excel 5.0. Duxbury Press, Belmont, CA.

- Winston, W.L. 2004. Microsoft Excel Data Analysis and Business Modeling. Microsoft Press, Redmond, WA.

Better Exponential Curve Fitting Using Excel

Mike Middleton, DSI 2010 San Diego

Michael R. Middleton, Ph.D.
Decision Toolworks
Mike@DecisionToolworks.com
415.310.7190



PowerPoint Slides, Slides PDF File, and Excel Workbook

http://www.DecisionToolworks.com/ExpCurveFit2010.pptx
http://www.DecisionToolworks.com/ExpCurveFit2010.pdf
http://www.DecisionToolworks.com/ExpCurveFit2010.xlsx