INCORRECT NONLINEAR CURVE FITS IN EXCEL

Rick Hesse, Graziadio School of Management, Pepperdine University
P. O. Box 6175, Ventura, CA 93006   805/644-5396   RickHesse@AOL.com

ABSTRACT

A big problem with using Excel for nonlinear curve fits, such as exponential or power fits, is that the exponential and power fit graph functions are done incorrectly. An example for both types of nonlinear fit is given, comparing the correct fit using the Solver included in Excel with the incorrect Excel curve fit.

INCORRECT FITS

In the "early" days of statistics, before calculators and computers, nonlinear calculations were so gruesome and abhorrent, that shortcuts were often sought to ease the pain of the calculations. Some statisticians thought that if you wanted an exponential fit of the data ($Y = aG^x$) or a power fit ($Y = aX^b$), taking the logarithms of the data and finding the linear fit and then converting back by putting the answers to the power "e" would give a correct nonlinear fit. This is a very serious mathematical mistake, because what minimizes the sum of the logarithms does not minimize the logarithm of the sum. This is a very serious error, about which I have warned in the past [1,2]. I was surprised when I recently investigated the automatic curve fit function in Excel for exponential and power fits and found them incorrectly done. What makes this error all the more serious is the ease with which users can simply click on the graph of the data and get an incorrect fit.

EXPONENTIAL CURVE FIT

Consider the sales figures for 12 years (not consecutive) in Figure 1, for which we want to fit an exponential curve of the form $Y = aG^x$, where G is the growth parameter and is sometimes written as e^b. The rate of growth, R, is G-1.

Year	1	3	4	6	7	8
Sales($K)	$300	$372	$423	$608	$721	$816
Year	9	10	11	12	14	15
Sales($K)	$920	$1,135	$1,315	$1,530	$2,172	$2,481

Figure 1. Sales Figures for Exponential Curve Fit

The Excel fit is determined by making an XY (Scatter) chart as shown in Figure 2.
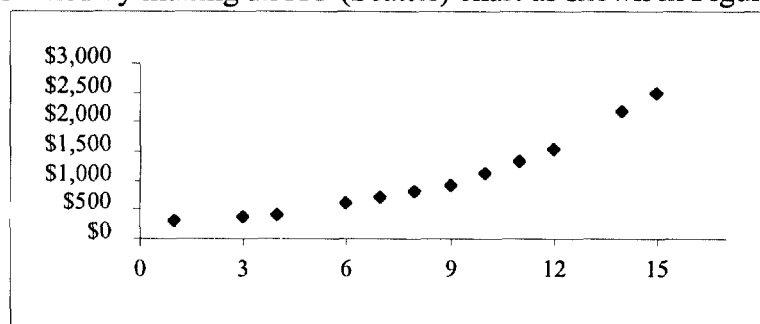

Figure 2. XY (Scatter) Chart for Data

Then a right click on the graph of the data points brings up the trendline menu, as shown in Figure 3. The tab for options can also be clicked and you may request the formula. Using this formula, the errors are determined on a spreadsheet, as well as RMSE, which is the square root of the sum of the errors divided by the degrees of freedom (12 points – two parameters = 10).
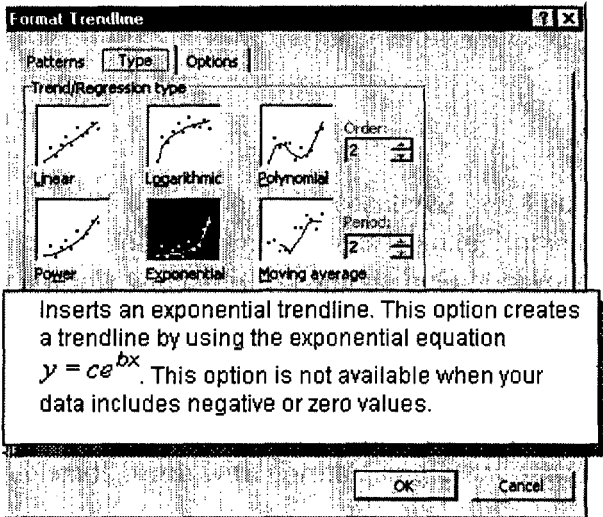

The trendline is given as $y = 238.9287*e^{0.1552*x}$ (which is $y = 238.9287*(1.1679)^x$) and the RMSE is 30.98. The graph in Figure 4 looks like a good fit, but then looks can be deceiving.
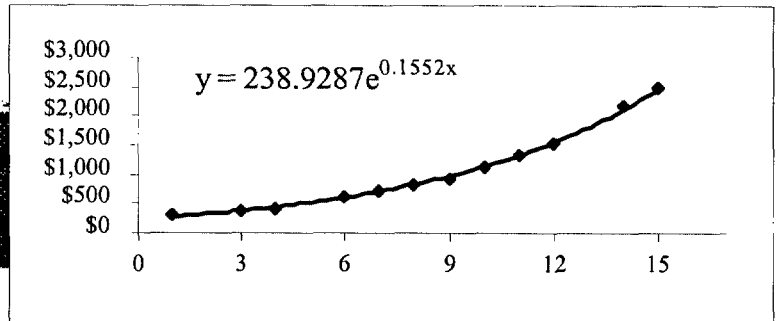

Figure 3. Excel Format Trendline Menu   **Figure 4.** Excel Trendline and Formula

SOLVER TRENDLINE FIT

The model to minimize the RMSE is found in cells A9:D22, with the data in columns A and B, the model forecast in column C and the signed error in column D. C6 contains the formula to compute the RMSE, which is =SQRT(SUMSQ(D11:D22)/F6) and in Figure 4 is equal to the RMSE for the Excel Trendline fit. The variables are the values of a (B5) and G (B6). This gives us an unconstrained nonlinear problem with a good starting point. The solver is invoked by calling Tools ... Solver from the Excel menu and the setup is shown in Figure 5.
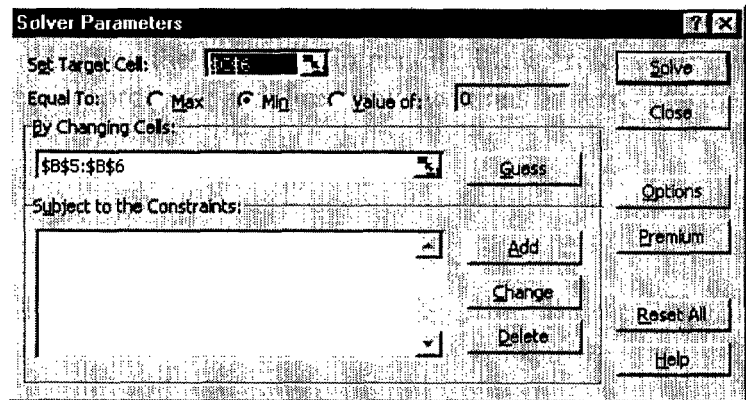

Figure 5. Solver Menu for Optimal Curve Fit

The spreadsheet in Figure 6 is set up for both the Excel Trendline fit and the Solver fit. To begin with, the values from the Excel Trendline fit are translated from the graph to the template in D5:D6 and pointed to in B5:B6 for the initial values of the Solver. The RMSE for using the Excel Trendline is larger by about 27.4% than the model using the Solver. The results are shown in Figure 6. The forecasts are also different, as well as the rate of growth (R = G-1).

	A	B	C	D	E	F
1	Nonlinear Exponential Regression Model Y = aG^x					
2	Exploding Business Sales					
3–5	Parameter	Solver Regression	Solver RMSE	Excel Regression	Excel RMSE	Degrees of Freedom
5	a = ▮▮▮		RMSE	238.92867	RMSE	Freedom
6	G = ▮▮▮	▮▮▮		1.1679	30.98	10
7		17.25%	Rate	16.79%		
8						
9–10	Y Sales (k)	X Year	Solver Forecast	Solver Error	Excel Forecast	Excel Error
11	▮	▮	$269.22	($30.78)	$279.05	($20.95)
12	▮	▮	$370.09	($1.91)	$380.65	$8.65
13	▮	▮	$433.92	$10.92	$444.58	$21.58
14	▮	▮	$596.50	($11.50)	$606.44	($1.56)
15	▮	▮	$699.38	($21.62)	$708.28	($12.72)
16	▮	▮	$820.00	$4.00	$827.23	$11.23
17	▮	▮	$961.42	$41.42	$966.16	$46.16
18	▮	▮	$1,127.23	($7.77)	$1,128.41	($6.59)
19	▮	▮	$1,321.64	$6.64	$1,317.91	$2.91
20	▮	▮	$1,549.58	$19.58	$1,539.24	$9.24
21	▮	▮	$2,130.18	($41.82)	$2,099.65	($72.35)
22	▮	▮	$2,497.56	$16.56	$2,452.26	($28.74)

Figure 6. Comparing Both Exponential Curve Fit Models

An even better fit results when the formula allows the Solver to find the best intercept instead of assuming c = 0 for the nonlinear equation $Y = aG^x + c$. Without going into details, the fit is a = 208.6496, G = 1.1790 and c = 38.0648 with an RMSE = 22.90, which makes the Excel model have 35.3% more RMSE than the 3-parameter Solver model. The growth rate again changes, this time up to 17.90% (versus 16.74% with Excel trendline and 17.25% with the 2-parameter Solver model).

POWER FIT TRENDLINE

This exact same problem of giving the incorrect fit occurs with the power fit trend line also; again probably because the logarithm of the data is taken, a straight line fit, and then the parameters raised to the exponential power. Better forecasting programs always do the nonlinear fit, rather than try to take the "quick & dirty" way out. Minimizing the actual errors is always preferred and even though statisticians will tell you that taking the logs of the data works when the data is "well behaved", you wouldn't be forecasting if your data were well behaved! When the data almost or exactly fits an exponential or power curve, taking the logs works pretty well – but the point is that you don't have data like that. (Of course an incorrect method works if there are little or no errors!). Figure 7 shows

the data for the power fit example, which is the hypothetical number of seconds it takes to compute the optimal solution to a problem given the number of variables in the model.

Nonlinear Power Regression Model $F = aX^b + c$

Computation Times

Parameters	Solver	Excel Regression	Excel	Degrees of
a =	RMSE	10.6457	RMSE	Freedom
b =		1.4359		13

Y Time (sec)	X Variables	Solver Forecast	Solver Error	Excel Forecast	Excel Error
		4.01	-5.99	10.65	0.65
		14.53	-27.47	28.80	-13.20
		30.86	-22.14	51.56	-1.44
		52.66	-13.34	77.92	11.92
		79.71	-18.29	107.36	9.36
		111.85	-32.15	139.48	-4.52
		148.94	17.94	174.04	43.04
		190.87	3.87	210.82	23.82
		237.56	-5.44	249.67	6.67
		288.92	58.92	290.45	60.45
		344.89	-24.11	333.05	-35.95
		405.40	48.40	377.37	20.37
		470.40	-49.60	423.34	-96.66
		539.83	15.83	470.87	-53.13
		613.66	-14.34	519.90	-108.10
		691.83		570.39	
		774.30		622.27	

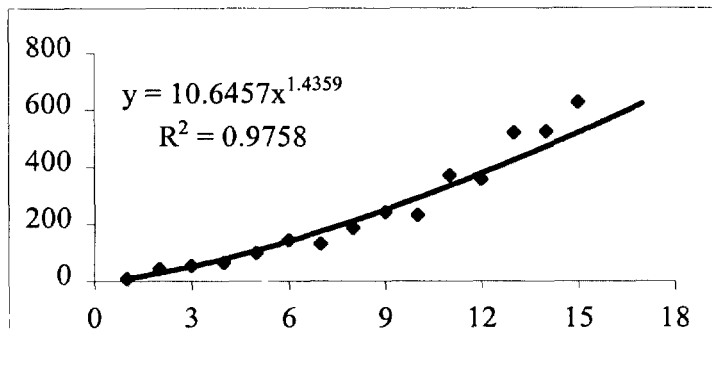$$y = 10.6457x^{1.4359}$$
$$R^2 = 0.9758$$

Figure 7. Optimal Solution for Solver Model Contrasted with Excel Trendline Solution

The same steps are taken with this example, the only difference being that the equation for the Power fit is $Y = aX^b$. Both the Solver optimized fit and the Excel quick & dirty fit are shown in Figure 7, and in this case, the Excel fit is 60.2% larger than the optimal value. If a third parameter, c, is allowed for a baseline, then the RMSE drops to 28.255 from 31.016, which means the Excel solution would then be 75.9% worse than the optimal curve fit ($y = 1.8715*X^{2.1277} + 14.7902$).

CONCLUSION

It is a simple matter to prove that Microsoft Excel is giving the incorrect nonlinear fits for both exponential and power fits. What is more difficult is convincing people not to use these fits or convincing Microsoft to do it correctly. The statistical part of Excel has always been a weak link (look at the atrocious DAT package) and this type of feature continues to degrade an otherwise fine product.

REFERENCES

[1] Hesse, Rick "Too Quick and Too Dirty: Least Squares for Exponential Curves," Decision Line, Vol. 14, No. 3, May 1983.

[2] Hesse, Rick "Son of Log Transformation or Return of the Living Dead," Decision Line, Vol. 18, No. 1, December 1986/January 1987.