Technical Tool Insight: Linear regression line

BY ACTIVE TRADER STAFF

FIGURE 1 FITTING A STRAIGHT LINE TO PRICES

A regression or "best-fit" line is calculated to minimize the difference between price points and the line. In doing so, the line approximates the slope (trend) of the prices.

Linear regression is a way to calculate a straight line that best fits a series of data points — that is, a line that most accurately reflects the slope, or trend, of that data. In terms of price analysis, a linear regression line is used to determine the trend of closing prices over a given time period. Because it is mathematically derived, a regression line, or "best-fit" line, is not based on subjective, visual analysis, as are standard trendlines.

Figure 1 (right) shows a group of five closing prices on a price chart. A straight line that goes through the "middle" of those five prices — a line for which the difference between it and each of the zigzagging prices is as small as possible — is a regression line.

Calculation

A regression line is calculated using the "least squares" method, which refers to finding the minimum squared (x*x, or x^2) differences between price points and a straight line. For example, if two closing prices are 2 and 3 points away (the distance being calculated vertically) from a straight line, the squared differences between the points and the line are 4 and 9, respectively.

Why are the squared differences used, instead of just the differences? Figure 1 shows that some differences are negative (for points below the line) and others are positive (for points above the line). This makes it necessary to square all the differences, creating all positive values and making it possible to calculate a formula for the straight line.

The best-fit line is the line for which the sum of the squared differences between each price and the straight line are minimized.

The formula for a straight line (*y*) is:

$$y = b + m*x$$

where

x = the "time" of the price (the x-axis value)

b = the initial value of the line when "x" is equal to zero (the "intercept" value — i.e., the point at which the line intercepts the vertical axis);

m = the slope of the line, which is the rate at which the line rises or falls. In other words, b is how much y changes for a one-unit change in x (e.g., .75 points per day).

As prices change, the slope of the line also changes. When a market is rising sharply the slope (b) has a high value and the line will be steep. As the market slows down, the slope value decreases and the line will slope upward more gently.

When calculating a straight line to *N* prices, the "best-fit" coefficients **b** and **m** can be solved for by:

$$b = [(4N + 2)/(N^2 - N)] \sum_{x=1}^{N} p(x) + [6/(N^2 - N)] \sum_{x=1}^{N} x*p(x)$$

$$m = [12/(N^3 - N)] \sum_{x=1}^{N} x*p(x) - [6/N^2\text{-}N)] \sum_{x=1}^{N} p(x).$$

where

p(x) is the price at point x.

N is the number of prices used to calculate the coefficients — e.g., N = 5 for a five-day regression calculation. In this case, the first day p(1) in the price series is 1 and the last price p(N) in the series is 5.

$\sum_{x=1}^{N} p(x)$ is the sum the prices for p(1) through p(N). For example, if N = 5 and the prices for days 1 and 2 are 10, 11, 12, 13 and 14, respectively, the sum is 60.

$\sum_{x=1}^{N} x*p(x)$ is the sum of the products of time (x) and price (p). For example, the products of the prices used in the

previous calculation are 10 (1*10), 22 (2*11), 36 (3*12), 52 (4*13) and 70 (5*14), and the sum of those products is 190.

Figure 2 (top) show the calculations and chart of three five-day regression lines calculated at different points over a 10-day period: Line A covers days 1 through 5; line B represents days 4 through 8; and line C is days 6 through 10. The linear regression estimates for the slopes (b) and intercept values (a) are listed in the third and fourth columns. The values for each of the five points that make up regression lines A, B and C are in the final three columns

The slope for line A (days 1-5), which accompanies an upward trend, is 0.60. Price continued to rally higher in days 6 through 8, but at a slower rate, which resulted in a slope of .43 for line B. For line C, when price moved sideways to lower, the slope was -0.07.(Although analysis programs automatically perform all these calculations and plot the resulting regression line, those interested in walking through the math for line A in Figure 2 can go to the Web Extra for this article at www.activetradermag.com/toc.htm between May 12 and May 30, 2003.)

Figure 3 (bottom) shows two regression lines, one covering the most recent 30 days, and another encompassing the most recent 90 days. Note the first line indicates uptrending prices over the shorter time frame and the second line reflects downtrending prices on the longer time frame.

Interpretation and application

Unlike standard "visual" trendlines, regression lines are not typically used to determine trend changes when price penetrates them to the upside or downside. As Figure 3 shows, price fluctuates above and below the regression lines, because the lines are calculated to come as near as possible to all the closing prices in the periods they cover.

A regression line shows the slope, or trend, over a given time period; the farther an individual price point is from the line, the more it has strayed from the typical price behavior for this period.

FIGURE 2 REGRESSION LINES

The following calculations resulted in the three different five-day regression lines on the chart below.

Day	Price	Slope (m)	Intercept (b)	Line A	Line B	Line C
1	21.25			20.50		
2	20.50			21.10		
3	21.00			21.70		
4	22.50			22.30	22.71	
5	23.25	0.60	19.90	22.90	23.14	
6	24.20				23.57	23.95
7	23.25				24.00	23.89
8	24.66	0.43	22.28		24.44	23.82
9	23.00					23.76
10	24.00	-0.07	24.02			23.69

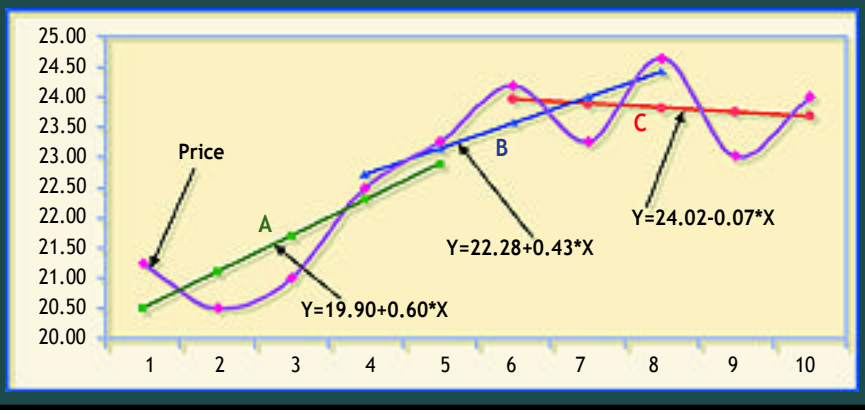

FIGURE 3 LONG-TERM SLOPE VS. SHORT-TERM SLOPE

The uptrending (red) line is a 30-day regression line, while the downtrending (blue) line is a 90-day regression line.

Source: TradeStation Platform by TradeStation Group