Linear Regression Projection of Periodicity

Charles D. Kirkpatrick II

The complex period reconnaissance methods, such *as* maximum entropy and spectral analysis, estimate the optimal cycle parameters existing in time-series data. As many technical analysts have long recognized, however, the results from these techniques rarely project with any accuracy the inherent periodicity in trading-market time series data. This is because market data is not affected by the combined constant amplitude and period presumed by these methods. Instead, market price data includes changing price volatility which negates the constant-amplitude inference. Indeed, any cyclical scanning method which presumes constant amplitude in market price data is suspicious.

In addition to amplitude volatility, trading market cycles also exhibit the curious habit of inversion whereby cyclical highs suddenly occur when cyclical lows are expected and vice versa. Inversion occasionally occurs when a sudden, dramatic, non-market event influences market psychology or when a larger cycle reversal is occurring. It is a phenomenon which traditional reconnaissance methods are unable to decipher and which upsets the averaging process of most cyclical scanning methods.

Avoiding the potential errors involved in spectral scaning methods, some analysts take the intervals between identifiable price cycle lows (or highs) and project periodicity rather from an average of these intervals. This follows in the tradition of J. M. Hurst's "Profit Magic of Stock Transaction Timing" (Prentice-Hall, Englewood Cliffs, NJ, 1970) and is an easy and useful though somewhat inaccurate procedure. Because it uses the intervals between points rather than low (or high) points themselves, this technique fails to establish an accurate base of points. It is dangerous because the last low (or high), which is often taken as the starting point for projections into the future, may be skewed to either side of the ideal low (high) point and thus may skew any projection by an equal and unknown amount. Additionally, because intervals rather than low (high) points are used, the averaging method cannot obtain the statistical advantages of other techniques. It is limited to the mathematics of averaging.

To avoid this averaging error, the inaccuracies associated with conventional cyclical scanning methods, and the problems associated with inversion, and to still have a statistical means of testing significance, I have found periodicity can be estimated and projected with relative ease through the technique of linear regression. Linear regression completely disregards and thus avoids the question of variable amplitude in cyclical behavior. Being one-dimensional (measuring time only, for example), rather than two-dimensional (time and amplitude), it can project along a time line, with quantifiable accuracy, the expected points at which periodic events (such as tops and bottoms) are likely to occur, thereby avoiding the projection problems associated with interval averaging. And finally, as well as providing a statistical means of quantifying error in projections, it can also account for inversions and other non-standard cyclical behavior (such as the time series changes in periodicity). In short, it is a simple but superior means of analyzing cyclical behavior in trading markets.

Linear Regression

Linear regression, sometimes called the "least squares method," estimates the straight-line relationship between an independent variable and a dependent variable in terms defined by the simple linear formula:

$$Y = a + b\,(x) \qquad (1)$$

where Y *is* the dependent variable (in this instance, a time or date); X is the independent variable (sequential cycle number); a is a constant (starting point for the first cycle), also known as the Y intercept; and b is a constant coefficient (time length of period), also known as the slope.

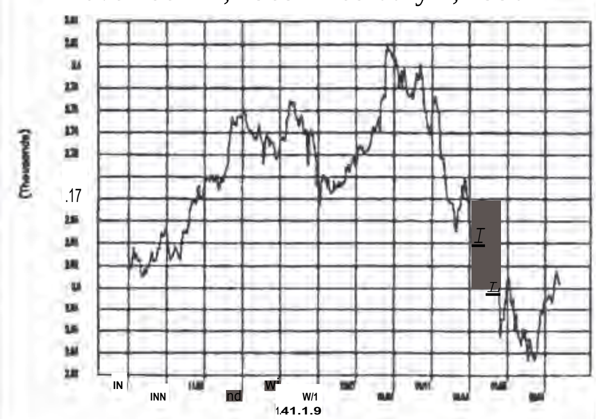
Linear regression has many other attributes which are not pertinent to this discussion. Any textbook on statistics will describe its nuances in more detail. Here we are more interested in its use rather than in its derivation.

An Example: Linear Regression Projection of 46-Hour Periodicity in the Dow Jones Industrial Average

Data Collection

The best way to understand the use of linear regression in market cycle projection is to follow an example. For this purpose, the Dow Jones Industrial hourly ("DJH") figures (10 am, 11 am, 12 noon, 1 pm, 2 pm, 3 pm and close) from November 12, 1989 to March 7, 1990 are used to demonstrate and later to project a specific market periodicity. These dates are used because they were the most recent at the time of the original writing of this paper. The results have continued to hold true through several revisions.

Chart A
HOURLY DOW JONES INDUSTRIAL
November 12, 1989 - February 2, 1990



The first step in any investigation of periodicity is to gather sufficient price data, generally enough to demonstrate eight to ten instances of a suspected cyclical period. The cycle period initially can be estimated through the use of a spectral scan—also called a "systematic period reconnaissance." This technique places time series data into a succession of arrays, each of which is added and averaged, and a sinusoid fitted to each average. The characteristics of the sinusoids are then compared to suggest which cycles are the most dominant. Obviously a computer and adequate software make this procedure easier. In the DJH data, a spectral scan (using the C.A.P.II program from The Foundation for the Study of Cycles, Irvine, CA) suggested a period of roughly 40 hour was the most dominant. Thus, 400 data points (40 hours times 10 cycles) were gathered, as a minimum. (An additional 150 data points were also gathered at the end of the main time series. These are used later in this example to test the accuracy of the linear regression method.)

The next step is to create and enter into a table *(Table A)* dates or times, converted into a sumber sequence (columns 1 and 2), and their corresponding prices (column 3). Dates and times themselves are not easily manipulated mathematically without conversion into a simple integral series. Dates, for example, must be adjusted for weekends and holidays, and hourly data must be adjusted for non-trading time. The usual manner in which dates and times are represented in time series is by attaching a sequential series of integers to each successive date or time. Thus January 3rd, 4th, and 7th (assuming a weekend in between) could be represented as date number 1, 2 and 3. 3 o'clock, 4 o'clock of January 3 and 10 o'clock of January 4 (a new day) could equally as well be represented by 1, 2, and 3 in an hourly sequence. The starting number is irrelevant, but the sequence must be iterative and the numbers equally spaced, i.e., 1, 2, 3, 4, etc., or 234, 235, 236, 237 , etc., not 1, 4, 3, 9, etc.

The next step is to determine what specific periodicity to project. In the example, a spectral scan implied a significant harmonic in the vicinity of 40 hours. Since this is a useful trading period, not too short to miss directional turns and not too long to make calculations cumbersome (generally between 20 and 60 intervals is ideal), a 41-hour, centered moving-average (always an odd number, one integer larger than the expected cycle period, to position the average on a mid-cycle date number) is calculated and positioned *as* a column (column 4) in the table (Table A), each centered moving-average price corresponding to its respective date number. Since a moving-average tends to dampen out the cyclical activity of period less than half the number of prices averaged (i.e., in the example, a 41-hour moving average dampens out cyclicality of less than 82 hours), it will reflect only the influence of cycles longer than twice its average (i.e., it will reflect the motion due to cycles longer than 82 hours).

We now have four columns in the table (Table A)—the date and time (column 1), the date number (column 2), the price (column 3), and the centered moving-average (column 4). One more column is needed (column 5). It is calculated from the third and fourth columns, namely a difference (or ratio) of the price (column 3) to the centered moving-average (column 4). Sometimes the price is also smoothed over three periods or a three-interval median is used to dampen the very short-term price oscillations and to reduce the effect of outrageous, one-time aberrations. Ideally then, the fifth calculated column should be a difference between the three-interval moving average of the price and the centered moving-average (a difference of the 3-hour moving average to the 41-hour moving average, in the example).

Since the three-interval price column represents all data with cyclical periods greater than 6 (twice the average number) and the centered moving average column represents all activity greater than 82 hours, the difference should represent all activity between 6 hours and 82 hours. The effects of periodicity longer than 82 hours and shorter than 6 hours is thereby reduced.

The completed table should look like this:

Table A
Example of Master Table

(1) DATE	(2) NO	(3) DJ HOUR	(4) 41HOUR CTR MOVAVG	(5)
11/28	70	2692.62	2686.63	8.55
	71	2697.95	2688.26	8.77
	72	2700.51	2690.18	9.13
	73	2699.45	2692.37	6.29
	74	2696.03	2694.62	4.68
	75	2702.43	2697.05	3.11
	76	2702.01	2698.99	2.66
11/29	77	2700.51	2700.84	-1.54

Another way to portray this data, where cyclicality can be seen more clearly, is to plot along a horizontal axis a bar or line chart representing the price deviations from the moving average (column 5 in Table A).

Initial Estimate of Deviations

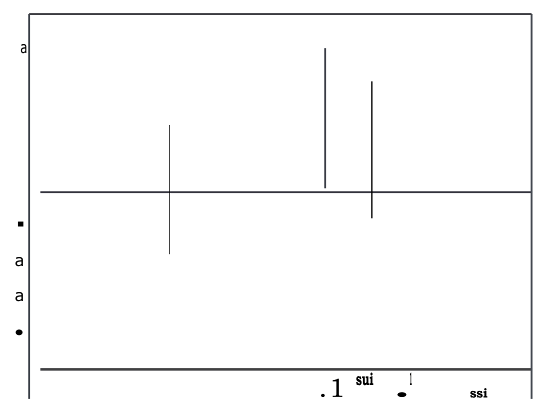
The periodic peaks and valleys can be seen clearly in *Chart B.* Estimates of periodicity can now be determined. Notice in the plot of the DJH deviations how the most clearly defined deviations seem to occur at lows rather than highs. This is very common in market price data.

The next step is to construct a second table—the cycle table *(Table* B)—which includes the cycle number (X) and the date number (Y) of each major bottom. There are several methods to estimate where significant lows have occurred. One is the use of a mechanical device called a "cycle finder;" another is the averaging of the interval lengths between obvious lows. In this technique, take a sheet of paper and mark perpendicular to the edge a long line. Place this line on the chart at the first obvious price or deviation low. Look on the chart to the right until you see the next best looking low at the approximate period (around 40 hours) distance from the first low and mark a short perpendicular line on the paper at this point and a dot on the chart at the price low. Best looking deviations are those with normal appearing distributions, i.e., one clear peak or valley and a symmetrical drop-off in deviation on either side. Now move the paper such that the long line lines up with the dot at the next price low and repeat the process to the next obvious low. After several lows have been identified, the paper will have the long line to the left and a series of shorter lines, hopeful. ly bunched together, representing cycle periods from low to low.

On the paper, now mark a star where the most number of short lines occurs. This is the point where the most dominant cycle seems to make lows when measured from the long line. Take the paper and put the long line again over the first obvious low and look for a low on the price or deviation chart (Chart B) closest to the star mark and record the date number of this low in the cycle table (Table B). Move the paper to the right again, putting the long line on the recorded low, look for a low in the vicinity of the star and record the date number. Repeat the process, each time recording the date number of each successive cyclical low.

Chart B

PRICE DEVIATIONS
Dow Jones Industrial Average Hourly
November 13, 1989 - February 2, 1990
Showing Three-hour Deviations from a 41-hour, Centered Moving-Average



A third method for determining the cycle lows is to look for large deviations in the master table. (Table A) at the expected trading cycle interval (40 hours in the example). This method is similar to the second method (using the paper edge method) only it uses the master table (Table A) instead of a plot. In this case, look for an early obvious low and record its date number in the cycle table (Table B). Now add the period estimate (40 hours in the example) and within a reasonable proximity to this date number look for another obvious large negative deviation in column 5. In the example, the first obvious low is at date number 46 (cycle number 1). The next expected low is in the vicinity of date number 86 (46 best low plus the 40 hour estimated period). A distinct deviation occurs at date number 87. This should be recorded as cycle number 2.

Next, add another estimated period of 40 hours to the last estimated point. One should always continue estimating from the original best looking deviation by the estimated intervals. In the example, the next date number to inspect for a low in its vicinity is 126 (86 plus 40). Here there is a very clear maximum deviation and price low at date number 136. This is recorded in the cycle table (Table B) as the next cyclical low (number 3). Since the actual cycles appear to be a little longer than the original 40-hour estimate, use a slightly higher number for the estimated interval—usually an average of the intervals so far is sufficient (136/3 = 45.3 hours). Continuing this process of adding (or subtracting) one estimated period (now 45 hours), a table (Table B) is constructed including all possible periodic deviations with their respective date number.

The table resulting from the DJH data is:

Table B
Cycle Table

Cycle Number X	Date Number
1	42
2	87
3	140
4	180
5	222
6	280
7	319
AVERAGE: 4.0	181.4

Sometimes a particular cycle low is not as well defined as others either because its deviation did not extend very far (remember, this technique disregards amplitude) or because some other cycle caused a series of smaller deviations, no one of which is the clear bottom. Each potential cycle low should be recorded separately for later testing, but only one should be placed in the cycle table (Table B) for each cycle number. Later, by watching how well the data fits the derived linear formula, the other lows will be substituted until the best regression fit is achieved.

Regression Calculation

The a and b constants in the linear formula (1) are derived from this table (Table B). Once the a and b constants have been derived, by substitution in the derived linear formula, any estimated X (estimated cycle number, X_5) will then project an estimated Y (estimated date number, Y_5). Since X in the table represents the cycle number, the ideal time estimate for each cycle low will be given by Y_E, the date number. Future Y_F's can be estimated by substituting future XE's.

First, the a and b constants are calculated by the following formulas:

$$a = (Y_A) - b\,(X_A)$$

$$b = S_{xy} / S_{xx}$$

where:

$$Y_A = \frac{Ey}{n}$$

$$X_A = \frac{Ex}{n}$$

$$S_{xy} = E(xy) - \left[\frac{E(x)\,E(y)}{n}\right]$$

$$S_{xx} = E(x^2) - \left[\frac{E(x)^2\,E(x)^2}{n}\right]$$

n = number of readings

This sounds more complicated than it is. Most spreadsheet programs for PC's have the formula already programmed whereby just the column of numbers need be substituted. If a spreadsheet program is not available, some hand calculators will perform the same calculations. Otherwise, provided the cycle table does not include too many cycles, the necessary calculations can be easily performed by hand.

For the DJH example:

$S_{xy} = 1299$

$S_{xx} = 28$

$n = 7$

and therefore

$a = -4.14285$

$b = 46.39286$

Substituting in (1):

$$Y_E \quad (-4.14285) + (46.39286)\, X_E \quad (9)$$

The estimate for cycle length (b) is 46.39286 hours and the initial low point occurred at date number (a) -4.14285. From this formula accurate identification of periodic lows and projections of future periodic lows can be made. For instance, we would expect the next low in Table B (number 8) to occur at date number 363 (-4.14285 + 46.39286 x 8).

Best Fit

In addition to the derived intercept and slope, we can determine another useful calculation called the "r squared" (r^2) which will help in finding the best fit. The r^2 is also called the "coefficient of determination." It is a complicated figure derived from the difference between what is predicted by the linear regression formula and what actually occurred. r2 is never greater than one nor less than zero. An r^2 of one indicates a perfect fit; zero indicates no fit at all. The closer r^2 is to one, the better. Thus, in any set of independent variables, the best fit will be that which results in an r^2 closest to one.

When attempting to identify the best price low for specific cycles, the r^2 can be very helpful. In the DJH example, in the vicinity of the expected cycle low at date number 319 in chart B (or Table A), for instance, there is also a large price deviation at date number 301. By substituting date number 301 in the cycle table (Table B) in place of 319 for cycle number 7 and solving the regression formula, the r^2 turns out to be 0.99635. When date number 319 is used, the r2 was 0.99872, somewhat closer to 1.00000. Since the latter is closer to one, the 319 date number is the best fit low. By this method any prospective deviation or series of deviations can be tested for best fit and substituted if it improves the r2.

Most spreadsheets and scientific calculators will figure r^2 from the data given in Table B. To hand calculate, the formula for r^2 is:

$$r \quad \frac{Sxy}{(Sxx)\,(Syy)} \qquad (10)$$

$$r^2 = r x r \qquad (11)$$

where

$$Syy = Ey2\ (Ev)^2 \qquad (12)$$

Again, this looks more complicated than it is.

Projection

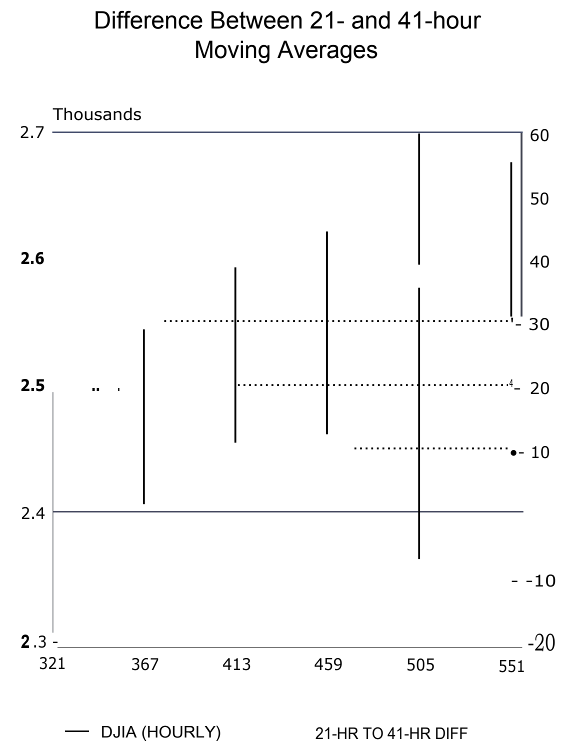
For projection into the future, to see when the next periodic deviation (low) is expected, a cycle number (X_E) should be substituted in the derived linear formula (3). In the DJH example, the last known X in the cycle table (Table B) is 7. When the next sequential numbers 8, 9, 10 and 11 are substituted for X_E in the derived formula (3), Y_E's of 367, 413, 459 and 506 result-the date numbers for the next expected cyclical lows.

Table C
CYCLE TABLE (continued)

Cycle No X	Date No Y	YE	YE)
1	42	42.2	-0.2
2	87	88.6	-1.6
3	140	135.0	5.0
4	180	181.4	-1.4
5	222	227.8	-5.8
.6	280	274.2	5.8
7	319	320.6	-1.6
8		367.0	
9		413.4	
10		459.8	
11		506.2	

To test the results of the example, the next 150 hours of prices was smoothed by the difference between the 21-hour and 41-hour centered moving-averages. These moving averages were selected so as to filter out the affects of cycles shorter than 42 hours (twice 21 hours) and longer than 82 hours (twice 41 hours) as per our earlier discussion on the construction of column 5 in Table B. This ratio and corresponding prices are plotted in *Chart C*. The chart is also marked with the projected date numbers for expected lows each approximately 46 hours apart as established from the derived formula. Happily, the actual lows came very close to when projected lows were expected (and have continued to do so through the time of this writing in June, 1990), lending credence to the linear regression technique. Notice in Chart C that the major lows near projected points 367 and 505 were of a higher order (the roughly 140-hour, or 20-day) cycle. The smaller cycles of 46-hours occur between these larger cycle lows.

Chart C
DOW JONES INDUSTRIAL AVERAGE
HOURLY
January 18, 1990 - March 7, 1990
Showing 46-Hour Projected Periodicity



Continuous Adjustment

Once a new target period low has been reached and clearly defined from column 5 in the master table (Table A), the new figures are again fitted to a newly regressed line using the same procedure as before. The new line should be very similar to the old line but different enough to account for the new data. In the example, the new linear formula is:

$$Y_E = (-3.892) + (46.310)\ X_E$$

The period (b) has shortened slightly to 46.310 hours from the earlier 46.393 hours. As time progresses, and future periodic lows are identified when they occur, the estimated starting point (a), the estimated periodicity (b) and location of expected lows (Y_E) will change. For example, *Table D* shows the a and b for the adjusted linear formula (3) in the four cycles following the original calculation:

Table D
Sequentially Adjusted Constant Values in Linear Formula
For Hourly Dow Jones Industrial Periodicity

No Points	Start a	Period b	Std Error	r2
7	-4.143	46.393	4.459	.9984
8	-3.892	46.310	4.082	.9989
9	-3.694	46.250	3.788	.9993
10	-1.133	45.551	5.097	.9987
11	-0.418	45.372	4.926	.9990

Error

Two columns can now be added to the cycle table (*Table C*). One column is the calculated Y_E for each successive cycle number X_E and the other is the difference between the actual date number Y and the Y_E, a gauge of the amount by which the projections vary from the specific observations (see Table C). A square root of the average square of these differences gives an estimate of the amount of error to be expected in future projections. This is also known as the "standard error of estimate" and when added and subtracted to any estimate should account for approximately 66 percent of all projections (two times the standard error of estimate should account for 90 percent of all projections and three times should account for 99 percent, assuming the errors are normally distributed).

In the DJH example, the standard error of estimate is +/– 4.459 hours, which implies that a cyclical low has a 66 percent chance of occurring between 41.934 and 50.852 hours (46.393+ /– 3.876 hours) after a previously projected low point. This calculation gives an analyst a range for the projected low point. More importantly, the standard error of estimate gives him a gauge by which he can measure his analysis. Occurrences of cycle deviations significantly outside the standard error of estimate suggest that something is wrong with the analysis and may be caused by a number of things: no cycle, a change in the cycle, incorrect prior identification of cyclicality, or inversion.

Inversion

Sometimes a Y_E will diverge widely from an actual Y. A sizable error, well outside the standard error of estimate, suggests either that earlier lows were incorrectly identified, or that an outside anomaly influenced the market at this particular moment which upset the prevailing rhythm and caused what is called an "inversion." When a sizable error occurs, an inversion should be suspected. An inversion occurs, for example, when highs suddenly appear at a time when lows are expected, or vice versa. If no obvious deviation occurs, periodicity likely has disappeared, or the current cycle is being dampened by another more dominant cycle. From this inversion point on, the linear regression model should use the high deviations rather than low deviations to measure periodicity. The same regression method will work, but it will project cyclical highs rather than lows. There were no apparent inversions in the DJH example.

Conclusion

Not finding standard spectral reconnaissance or other cycle scanning methods reliable in projecting market periodicity, I have used a simple method for projecting periodicity based on the fitting of observed cyclicality to a linear formula using regression techniques. One obvious pitfall to this methodology is that the initial estimates of cyclical behavior are subjectively determined. However, once the cyclical extreme points have been established, the linear regression model adjusts and gives an accurate account of periodicity, useful enough for projection. It even includes the ability to identify cycle inversion. The method is adaptable to all trading price data regardless of time interval and can be optimized by adjusting subjective observations to a calculated best fit. This method is another easily-used tool with valuable results for the trading-cycle advocate.

Charles 13. Kirkpatrick II (AB, Harvard, MBA, Wharton School, Unit,: Penn.) is President of Kirkpatrick & Company, Inc, an institutional investment advisory firm.