



### *Using The Fortune Chart*

# Predicting Market Data Using The Kalman Filter

*The Kalman filter is a two-stage algorithm that assumes there is a smooth trendline within the data that represents the true value of the market before being perturbed by market noise. Can this filter be used to forecast stock price movements?*

**by Rick Martinelli and Neil Rhoads**



Figure 1 shows daily opens for one year (252 days) of Ford Motor Co. (F). According to modern financial engineering principles, market data such as this is considered to be Brownian motion, which means that the daily price changes form a white-noise process. White noise is a random process in which consecutive

values are independent of each other, among other things, which means that each day, a price increase is just as likely as a decrease. In reality, it is not uncommon for a particular market item to have several consecutive down days or up days over a short time span. During such times, the prices are said to be *correlated*. The objective is to harness these correlations with a Kalman filter so you can forecast price movements.

In a 2006 article for STOCKS & COMMODITIES, a simple linear extrapolation was employed to predict tomorrow's price change. The prediction was then used to calculate the alpha statistic, which compares the predicted price change

KEN SMITH

to a recent average of price changes. Relatively large, positive values of alpha indicate a long position, and relatively large negative values indicate a short one. This procedure was backtested on a random selection of stocks and indexes to test its effectiveness. The indicated positions were taken, closed out the next trading day, and all profits and losses were accumulated in a chart called the *Fortune*. Surprisingly, of the 28 items tested, 20 produced greater profits than a simple buy & hold position for the same time period.

In this article, we will expand on the previous work, replacing the simple one-day predictor with a Kalman filter. The Kalman, as applied here, is a two-stage algorithm that assumes there is a smooth trendline within the data that represents the true value of the market item before being perturbed by market noise. In the first stage, a few previous trendline values are fit to a suitable model. It is then extrapolated to the next time value to generate a prediction and its error variance.

In the second stage, the corresponding data value is read and a new trend value is computed as a compromise between the prediction and actual data value. The compromise is based on the relative amounts of noise in the data and predictions. The filter then repeats the cycle of prediction and correction as each new data value is read.

The predicted price change and its standard deviation from the filter's first stage are combined to produce the alpha statistic, which is used to determine buy/sell signals. A simulated trading scheme executes those signals and, as before, profits and/or losses are accumulated in the *Fortune*. Items are selected for simulation based on a new data property called the *available profit*. The ratio of the fortune to available profit defines a characteristic of the filter called its *efficiency*.

The next sections discuss the Kalman filter and details of the simulation method. The last sections discuss results of simulations on real stock data and offer some conclusions.

## THE KALMAN FILTER

The Kalman filter is a recursive algorithm invented in the 1960s to track a moving target from noisy measurements of its position, and predict its future position. Applying this technology to financial market data, the noisy measurements become the sequence of prices

$$Y_1, Y_2, \dots, Y_N$$

where  $y_k$  is the price of a market item on day  $k$  and  $N$  is the total number of days; for the Ford data in Figure 1,  $N = 252$ . The simple equation

$$y_k = x_k + e_k$$

then says the each data value  $y_k$  is composed of a smooth trend  $x_k$  and a random noise component  $e_k$ , which is assumed to have zero mean. The objective here is to estimate the trend and use it to predict future data values. To do this, the Kal-

man filter employs an auxiliary equation to predict a future trend value from previous trend values. We chose

$$x_{k|k-1} = 3(x_{k-1} - x_{k-2}) + x_{k-3}$$

where  $x_{k|k-1}$  is the *predicted trend* value on day  $k$ , given data through day  $k-1$ , and the  $x_k$ 's are the previous filtered trend values. This is the *model*, or process by which trend values are predicted. The resulting filter is called the *quadratic filter*, because it assumes that every four consecutive trend values fit a quadratic curve. Associated with the model is a *process noise* sequence  $q_k$ , which in this case represents a correction term in the quadratic assumption.

A final trend estimate is made after the current data value  $y_k$  is input to the system

$$x_{k|k} = x_{k|k-1} + G_k(y_k - x_{k|k-1})$$

is called the *filtered trend* value on day  $k$ . The difference between actual and filtered values on day  $k$  is called the *residual*:  $r_k = y_k - x_{k|k}$ . The *innovation*  $v_k = y_k - x_{k|k-1}$  (in parentheses in the equation above) represents the new information in  $y_k$  not available in  $x_{k|k-1}$ . The *filter gain*  $G_k$  determines the contribution made by the innovation in the final estimate and thus controls the tradeoff between adherence to the model and fidelity to the data. The gain, in turn, is determined by the relative amounts of noise in the data and in the model, which are maintained in the filter.

The amount of noise in the data is denoted by  $R$  and is the variance of the residual sequence  $r_k$ . The amount of noise in the model, or process, is denoted by  $Q$  and is the variance of the process noise  $q_k$ . While  $R$  may be calculated from the residuals,  $Q$  is completely unknown. To estimate  $Q$ , we define the *tracking parameter*  $T$  by the equation

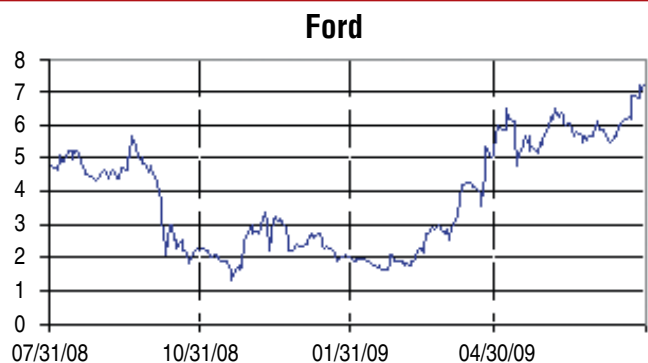
$$T = -\log(Q/R)$$

$T$  will be positive whenever  $Q < R$  and negative when  $Q > R$ . A negative  $T$  indicates the model has more noise than the original data and may need to be changed. In the following simulations, values of  $T$  ranging between -5 and 5 are tested and the value that minimizes the variance of the innovation sequence  $v_k$  is taken as the optimal  $T$  value.

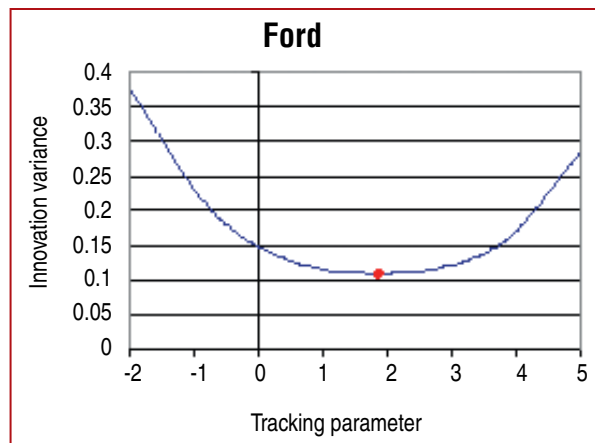
The data in Figure 1 was tracked with the quadratic filter and an optimal  $T = 1.86$  was determined. This corresponds to measurement noise that is about 72 times greater than model noise. Figure 2 shows how the innovation variance behaves as  $T$  varies and shows the minimum value at 1.86. The data was tracked again using optimal  $T$  and the resulting trend (a filtered version of Figure 1) is shown in Figure 3.

The gain for the track is shown in Figure 4; that it approaches a constant value (about 0.37) is typical of a well-modeled system. How well the filter and its intrinsic model perform depends on the filter's intended application. The current application is a one-day prediction of financial market price changes. To evaluate the filter in this context, the following section describes an idealized trading scheme

**It is not uncommon for a market item to have several consecutive down or up days over a short time.**



**FIGURE 1: OPEN PRICES.** On this daily chart of Ford Motor Co. (F) you can see the random nature of price movements.



**FIGURE 2: INNOVATION VARIANCE VS. TRACKING PARAMETER.** Here you see how the innovation variance behaves as T varies and shows the minimum value at 1.86.

where the predictions  $x_{k|k-1}$  and their standard deviation are used to calculate buy/sell signals.

### THE ALPHA INDICATOR

The Kalman filter provides predictions for each day in the data range (except the first few startup points). Figure 5 shows predictions for a short portion of the data in Figure 1 (green triangles). The filter also provides standard deviations of these predictions. If you let  $\Delta y_k = x_{k|k-1} - y_{k-1}$  denote the predicted price change on day k, and  $\sigma_k$  the standard de-

viation of that prediction, the alpha indicator for day k is defined as

$$\alpha_k = \Delta y_k / \sigma_k$$

Figure 6 shows the alpha sequence for the data in Figure 1, where the horizontal lines indicate the two alpha cutoff values. The irregular cyclic component that you see in alpha is typical of Kalman tracks of market data and is probably because of inherent correlations in the data. If alpha is positive, then the predicted movement is upward, and if alpha is greater than one, the predicted change exceeds the average. Conversely, if alpha is negative, the predicted movement is downward, and if alpha is less than -1, the predicted change once again exceeds the average. In the 2006 article,  $\alpha_k > 1$  indicated a long position and  $\alpha_k < -1$  a short position.

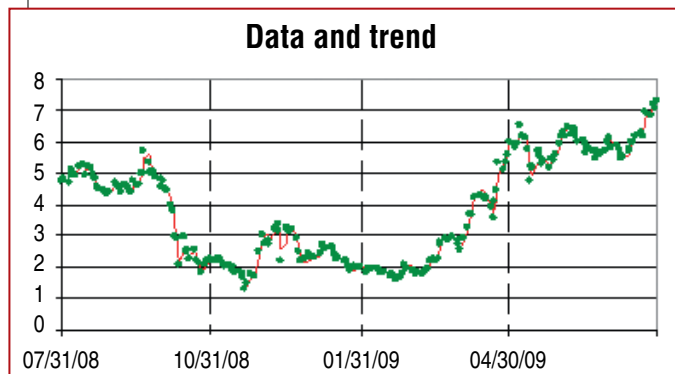
This article generalizes the situation. Here, a critical value C, called the *alpha cutoff*, is determined for each track, where  $\alpha_k > C$  indicates a long position,  $\alpha_k < -C$  indicates a short position, and no position is taken otherwise. Positions are canceled the following day and the associated profit/loss is accumulated in the fortune. The cutoff C is determined by testing values between zero and 3 in small increments, and choosing values that either yield the largest *last-day-fortune*, or the Fortune chart closer to the available profit line than any other value. These techniques will be discussed in part 2 of this series.

*Rick Martinelli and Neil Rhoads are with Haiku Laboratories.*

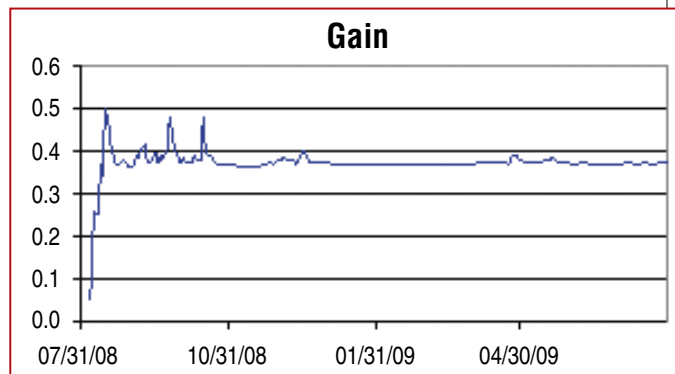
### SUGGESTED READING

Martinelli, Rick [2006]. "Harnessing The (Mis)Behavior Of Markets," *Technical Analysis of STOCKS & COMMODITIES*, Volume 24: June.

\_\_\_\_ [2008]. "Linear Estimation And The Kalman Filter," Haiku Laboratories.

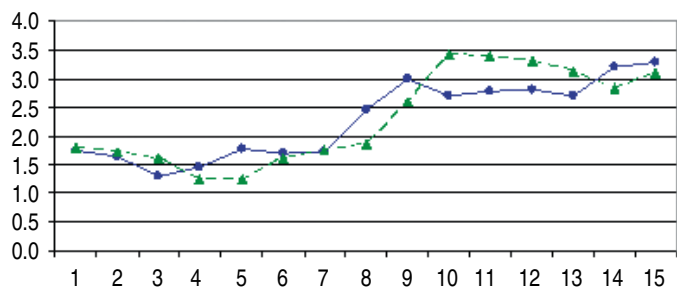


**FIGURE 3: THE KALMAN ESTIMATED TRENDLINE.** The data was tracked using tracking parameter ( $T = 1.86$ ) and the resulting trend is shown in red.



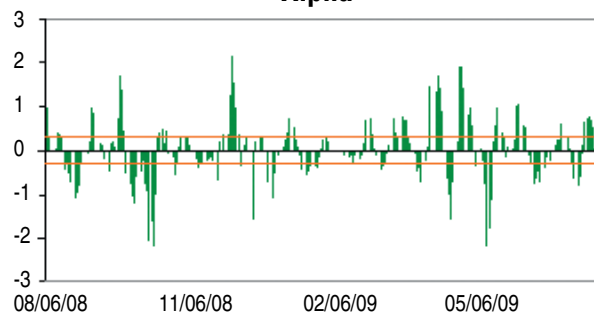
**FIGURE 4: THE KALMAN GAIN.** Note that it approaches a constant value of approximately 0.37. This is quite typical of a well-modeled system.

### Data and predictions



**FIGURE 5: KALMAN PREDICTIONS.** Here you see the Kalman predictions for a portion of the data from 11/18/08 to 12/09/08 (green) together with the data and the predictions for a short position of the data in Figure 1 (green triangles). The filter also provides standard deviations of these predictions.

### Alpha



**FIGURE 6: ALPHA VALUES FOR THE FORD DATA.** The horizontal lines indicate the alpha cutoff values -0.38 and +0.38.

**S&C**