KEN SMITH

The Future And The Filter

Predicting Market Data Using The Kalman Filter

Part 2

Can the Kalman filter be used to predict future price movement? In this second part of this series we answer this question.

by Rick Martinelli and Neil Rhoads

Previously, we discussed the Kalman filter and the alpha indicator. This time, we study the accumulation of profit/loss through the fortune chart. We also backtest the filter and analyze the results. The profit/loss on day k may be written as

$$P_k = A\,W_k\,(y_k/y_{k-1} - 1)$$

where the quantity in parentheses is the relative price-change, or *return*, on day k, A is the trade amount in dollars, and $W_k = 1$ if $\alpha_k > C$, $W_k = -1$ if $\alpha_k < -C$ and W_k = zero otherwise (W for wager). Note that $W_k = 0$ corresponds to no trade on day k and so P_k = zero as well. Further, note that A is the same for every trade. For the Ford data, C was found to be 0.38, and Figure 1 shows its P_k values, where the trade amount A was set to \$1.00 (zero values are not shown). The red point at November 28, 2008, represents the trade having the largest

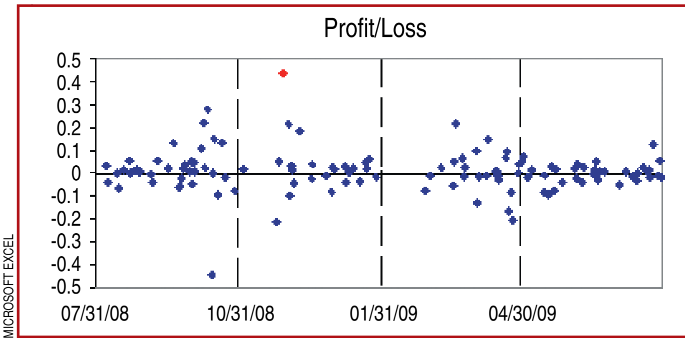


FIGURE 1: PROFIT/LOSS. Here you see the P_k values for the Ford data from Figure 1 in Part 1: T = 1.86, C = 0.36, A = 1. The red point at 11/28/08 represents the trade having the largest profit.

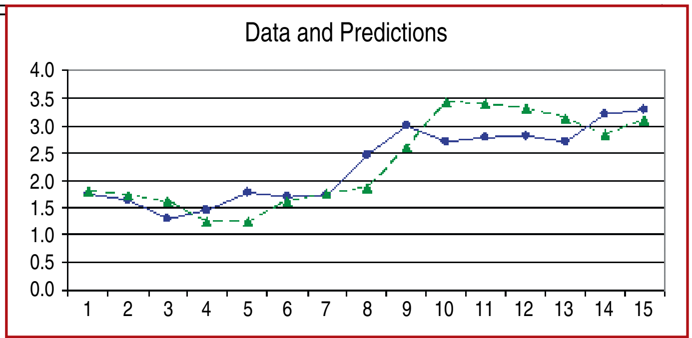


FIGURE 2: DATA AND PREDICTIONS. Here you see the Kalman predictions for a portion of the data from 11/18/08–12/09/09 (green) together with the data. The red point in Figure 1 corresponds to point 8 in Figure 2, Part 1.

profit. The reason for the large profit can be seen in Figure 2, Part 1, where the red point in Figure 1 corresponds to point 8 in Figure 2. The prediction of 1.84 was nowhere near the actual at 2.47, but it was in the *correct direction*, up from 1.72, resulting in a 0.436 profit. The fortune sequence is the accumulation of the P_ks:

$$F_n = \sum_{k=1}^{n} P_k$$

Its last value, F_N, is called the *last-day fortune* (LDF) and represents the amount of profit/loss realized at the end of the simulation. A graph of F_n versus day n, corresponding to the P_k data in Figure 1, is shown in Figure 3. The LDF is 1.37 on 124 trades, or about 1.1% average profit per trade. This is, of course, an idealized fortune in which there are no trade commissions, and trades can be transacted at the prescribed buy/sell prices (no slippage). It is used here primarily to evaluate the Kalman filter's ability to predict the direction a stock price will take.

The fortune plot is one indicator of that ability, and so is the *profit ratio*, defined as the ratio of number of profitable trades to total trades. In the Ford simulation that we discussed in Part 1, it was 0.59. The profit ratio can be roughly visualized as the ratio of number of points above the zero line to total points.

EFFICIENCY

A third way to evaluate the filter is by its *efficiency*. Suppose the scheme was able to capture *every* price change in the data — that is, correctly predict direction each trading day. Assuming A=1, the sum

$$S_n = \sum_{k=2}^{n} |y_k \ / \ y_{k-1} - 1|$$

of the absolute values of the returns represents the maximum amount of profit that can be realized by day n (using this scheme) and is called the *available profit* (AP) on day n. The last value, S_N, is the total AP in the data, and the ratio F_N/S_N is taken as the filter's efficiency. For the Ford example, S_N is 13.15 making the efficiency 0.104, or about 10%. Figure 4 plots F_n, the evolution of the fortune (green line) compared with S_n, the evolution of the AP, providing a view of the filter's efficiency over time. The available profit line has a desirable property, called *zero downside volatility,* which we would like to see in the fortune. Hence, we should choose the fortune line that is nearest the AP line, where the distance D between the two lines is calculated as:

$$D^2 = \sum_{k=1}^{n} (S_k - F_k)^2$$

The value of C that minimizes D can be taken as optimal C, which is the method used with the Ford data. The other method, maximizing the LDF, usually yields the same C

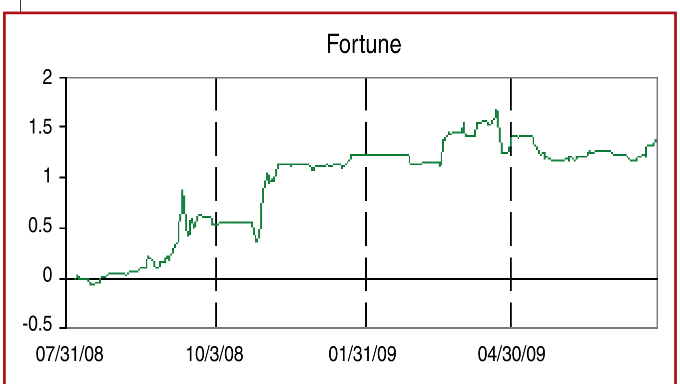


FIGURE 3: FORTUNE CHART. Here you see the fortune chart for the Ford data in Figure 1, Part 1. It is a quadratic model where T = 1.86, C = 0.38, A = 1, LDF = 1.37.

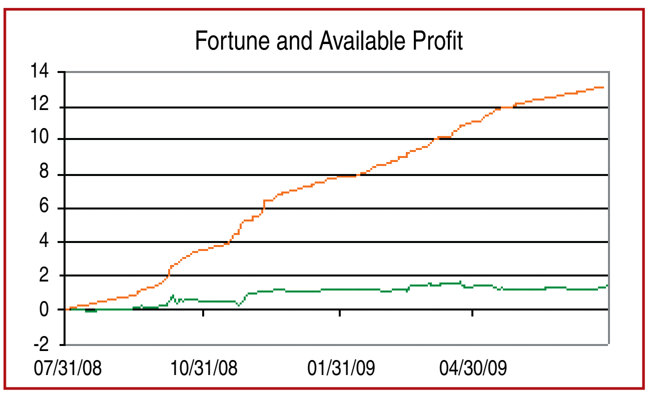


FIGURE 4: FORTUNE AND AVAILABLE PROFIT. Here you see a comparison of the fortune chart (green) with the Available Profit (AP) line for the Ford data.

	Symbol	End Date	Last Price	AP	T	Optimal C	LDF	Efficiency	Profit Ratio	# Trades	Dollar Return
1	MECA	10/30/08	1.62	39.15	4.720	0.08	1.901	4.9%	0.489	221	8.60
2	BUTL	07/13/09	0.02	32.92	3.380	0.06	5.926	18.0%	0.354	206	28.77
3	CTIC	08/18/09	1.54	31.30	-0.660	0.02	1.610	5.1%	0.390	187	8.61
4	ANPI	07/10/09	1.41	27.59	4.000	0.12	6.441	23.3%	0.521	192	33.55
5	BJCT	07/13/09	0.20	27.02	3.280	0.04	5.048	18.7%	0.358	232	21.76
6	BPUR	07/10/09	0.16	25.12	3.380	0.26	2.452	9.8%	0.493	150	16.35
7	TKO	07/29/09	0.14	23.87	3.540	0.36	1.653	6.9%	0.443	115	14.38
8	BBGI	07/10/09	2.23	22.03	2.020	0.34	3.176	14.4%	0.509	116	27.38
9	CENX	08/18/09	11.00	20.40	1.460	0.64	1.003	4.9%	0.632	38	26.39
10	RAE	07/29/09	1.75	20.34	3.140	0.36	1.836	9.0%	0.515	130	14.12
11	MDH	07/29/09	1.26	19.54	4.120	0.18	5.536	28.3%	0.555	182	30.42
12	ACAS	07/29/09	3.35	19.47	0.640	0.06	1.787	9.2%	0.500	192	9.31
13	HEB	07/29/09	2.30	18.86	3.940	2.38	1.589	8.4%	0.611	18	88.28
14	CLZR	08/18/09	1.38	18.64	1.940	0.24	1.798	9.6%	0.467	120	14.98
15	NG	07/29/09	3.94	18.22	1.480	0.06	4.907	26.9%	0.563	222	22.10
16	GAN	07/29/09	14.24	18.16	3.300	0.40	1.456	8.0%	0.451	122	11.93
17	PSTI	08/20/09	1.63	17.04	2.000	0.22	1.343	7.9%	0.425	160	8.39
18	BASI	07/10/09	0.85	16.97	4.800	0.00	8.898	52.4%	0.470	247	36.02
19	PAL	07/29/09	2.54	16.93	1.500	0.08	1.327	7.8%	0.507	217	6.11
20	BCRX	07/13/09	3.86	16.68	1.600	0.40	0.463	2.8%	0.538	106	4.37
21	TIV	07/29/09	0.99	16.53	1.840	0.00	1.306	7.9%	0.498	247	5.29
22	BANR	07/10/09	2.93	16.31	1.960	0.32	1.522	9.3%	0.579	95	16.02
23	ABG	07/29/09	11.99	16.27	1.720	1.04	1.116	6.9%	0.611	36	30.99
24	FRG	07/29/09	3.68	15.70	1.540	0.16	1.316	8.4%	0.585	200	6.58
25	CNXT	08/18/09	2.35	14.61	0.860	0.16	2.201	15.1%	0.525	160	13.76
26	BPFH	07/13/09	4.48	13.92	1.280	0.74	0.122	0.9%	0.514	37	3.29
27	FAC	06/10/09	2.11	13.72	2.540	0.04	1.368	10.0%	0.465	228	6.00
28	BPOP	07/13/09	1.40	13.16	1.560	0.10	1.829	13.9%	0.543	173	10.57
29	F	07/29/09	7.08	13.15	1.860	0.38	1.367	10.4%	0.589	124	11.02
30	BELFA	07/10/09	13.39	13.12	2.920	0.28	3.131	23.9%	0.546	152	20.60
31	CACH	08/18/09	4.43	12.82	1.820	0.18	1.160	9.1%	0.522	113	10.26
32	PWR	06/10/09	25.66	9.76	1.800	0.12	0.105	1.1%	0.505	206	0.51
33	CMCO	08/18/09	14.03	9.60	1.580	0.12	1.019	10.6%	0.539	206	4.95
34	CAKE	08/18/09	17.91	9.18	2.600	0.94	0.489	5.3%	0.545	33	14.82
35	GE	06/11/09	13.43	8.38	2.740	0.66	0.717	8.6%	0.614	57	12.58
36	CALM	08/18/09	28.72	7.50	1.180	0.28	0.941	12.5%	0.532	126	7.47

FIGURE 5: RESULTS OF SIMULATION OF 36 LARGE AP STOCKS. The results for the Ford data are shown. The last column is the average profit/loss in dollars, or dollar return, based on a trade amount of $1,000.

value, but not always. Given a choice, we take the largest C value because although the LDF may be smaller, the fortune line is smoother.

BACKTESTING THE FILTER

The filter was tested on one year of daily opens for a large group of selected stocks (all data obtained from Yahoo Finance). The simulations involved two optimizations: The first optimization determines the best Kalman tracking parameter and the second finds the best alpha cutoff. The simulation proceeds in five steps:

1. Track one year of data multiple times to find the optimal tracking parameter T
2. Track the data once more using optimal T
3. Use the resulting Kalman predictions and their standard deviations to find the alpha sequence α_k
4. Use that alpha sequence to find the optimal cutoff C
5. Use the optimal cutoff C to calculate the fortune chart.

While we chose the Ford data more or less at random, the items used in the simulations were carefully selected to have the largest available profit. We scanned opening prices for approximately 5,000 stocks for their one-year available profit and many of the stocks having APs greater than seven were tracked with the quadratic filter. One reason for choosing opens is that they usually have the largest AP of the four daily prices. A representative sample of the results is shown in Figure 5 sorted by AP. We have included the results for the Ford data and highlighted them.

The last column is the average profit/loss in dollars, or dollar return, based on a trade amount A = $1,000. To get a better feel for the nature of these results, we plotted key features separately, starting with the sorted APs in Figure 6.

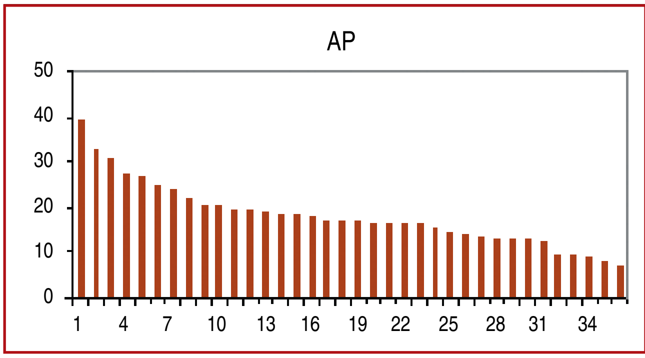


FIGURE 6: SORTED AVAILABLE PROFITS. Here you see the sorted APs for the 36 items in the table in Figure 5.

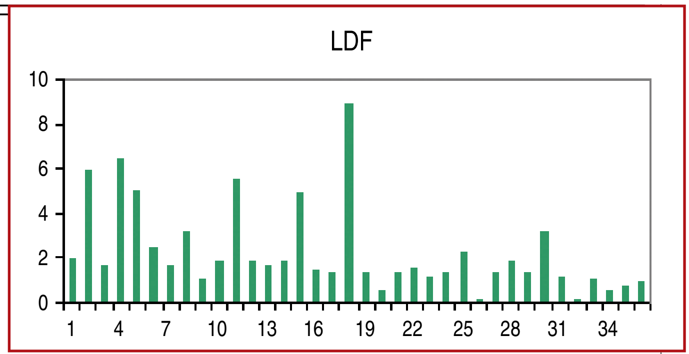


FIGURE 7: LAST DAY FORTUNE. Here you see the LDFs corresponding to the AP's in Figure 6.

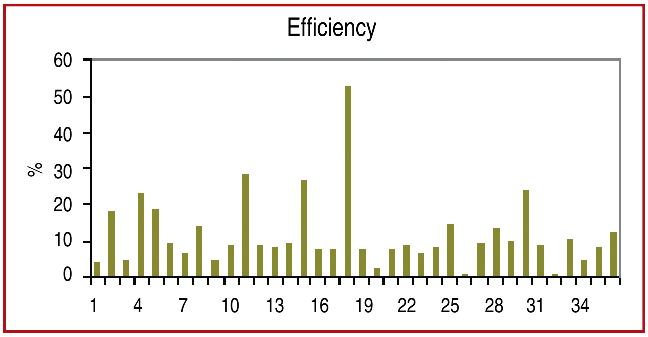


FIGURE 8: FILTER EFFICIENCY. Here you see the filter efficiencies corresponding to the AP's in Figure 6.

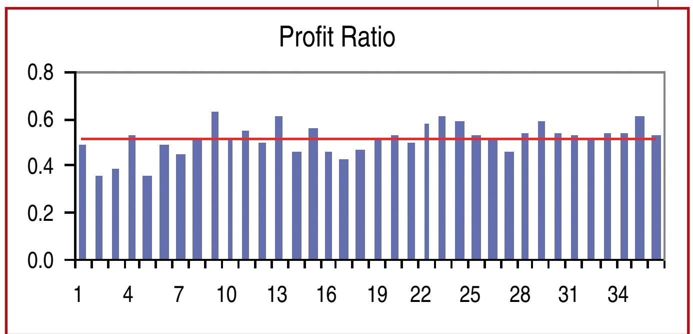


FIGURE 9: THE FILTER'S PROFIT RATIO. Here you see the profit ratios for the 36 test items and their average line (red).

Figures 6 and 7 show that filter efficiency is essentially uncorrelated with AP (actual correlation is 0.07). The obvious similarity between Figures 7 and 8 is artificial, because LDF values are the product of AP and efficiency. The average LDF is 2.2±1.9, a fairly wide range, and the average efficiency is 12.0%±9.5%, a slightly narrower range than the LDFs, but both are still dependent on the data. Available profit is a property of the data only and is independent of the filter. Only the filter's profit ratio (Figure 9) seems to be (nearly) independent of the data being tracked; its average is 0.51±0.07.

The highest efficiency is for no. 18 in the table (BASI) at an incredible 52%, with an accompanying LDF of 8.9. Data and fortune for BASI are shown in Figures 10A and B. BASI has the highest optimal *T* at 4.8, implying the quadratic model is well-suited to this data. BASI also has the smallest optimal *C* at 0.0, meaning every possible trade is executed. The least efficient is PWR at 1.1% with an LDF of 0.105, shown in Figures 11A and B. Its optimal *T* and *C* values, AP, and profit ratio values are all near average, suggesting that something inherent in this data may be responsible for the filter's relatively poor performance.

The stock having the fewest trades and best dollar return is HEB at about $88 per trade on 18 trades, shown in Figures 12A and B. Its fortune chart differs from the others in that there are relatively long periods of time where no trades occur. This is due to the large optimal *C* value at 2.38, the largest in the table. Fortune results like HEB are attractive for their minimal downward movement.

TCIC has the smallest tracking parameter at T = -0.66, meaning model noise is about double-data noise, suggesting another model should be considered.

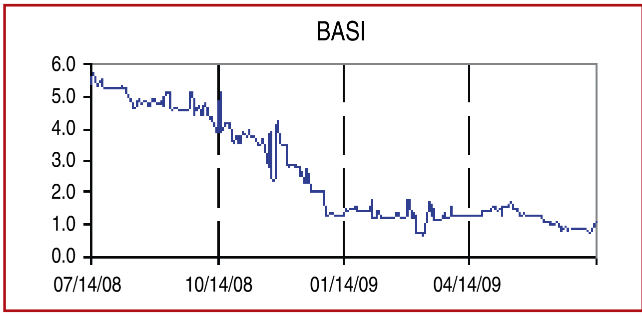
FORECASTING PRICE MOVEMENTS

The goal here was to determine if a Kalman filter could be exploited to predict the direction of stock price movements. Based on the LDFs of the selected stocks in the table, the implementation described here is effective in some cases (for example, BASI). The filter's effectiveness is traced to two primary factors: the nature of the data — namely, its available profit and amount of correlation — and the efficiency of the filter, which is determined by the filter's model in combination with a particular dataset. Only the profit ratio appears to be (nearly) independent of the data.

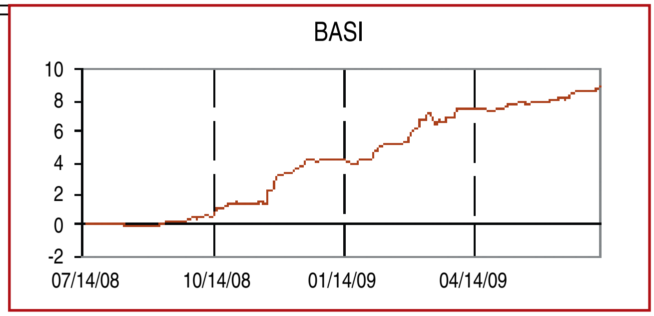
Selecting only the top performers from the table would lead to a highly profitable portfolio. Unfortunately, the simulation presented here could never be implemented in the real world because it requires all of the data on day 1 to determine the optimal C and T values responsible for the filter's performance.

However, application of this scheme to a shorter time period, say the previous quarter or two, should yield the best values for tomorrow's prediction. A simulation of this scheme must wait for another article.

Rick Martinelli and Neil Rhoads are with Haiku Laboratories. They may be contacted at info@haikulabs.com.

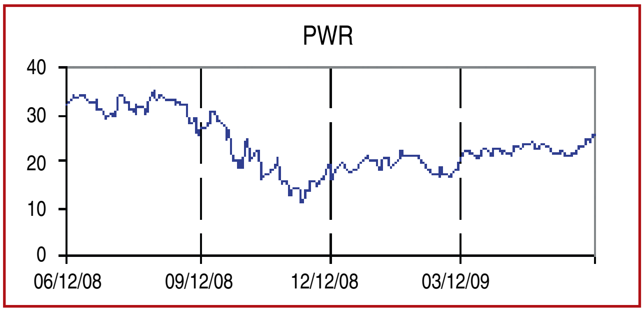


10A

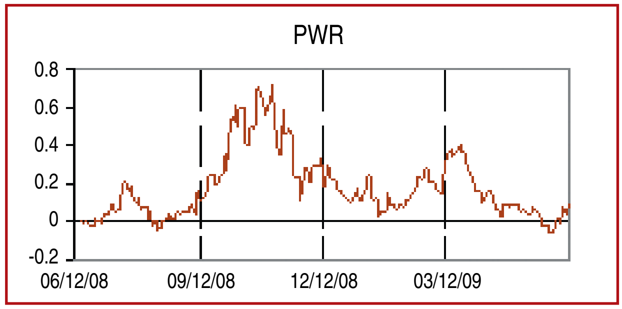


10B

FIGURE 10A and B: DATA AND FORTUNE FOR BASI ENDING JULY 10, 2009. BASI has the highest optimal T at 4.80, implying the quadratic model is well-suited to this data. It also has the smallest optimal C at zero, meaning every possible trade is executed. The AP=17.0, LDF=8.9, profit ratio 0.47, 247 trades, and is about $36 per trade.

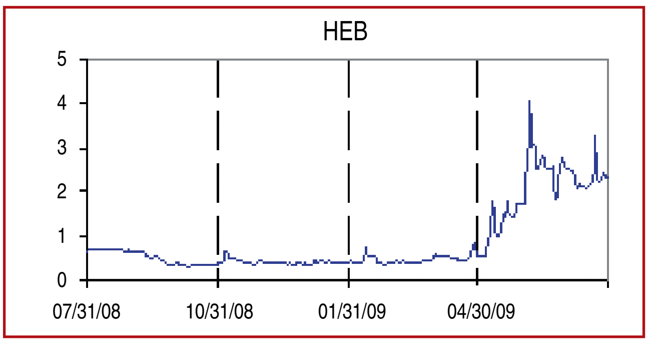


11A

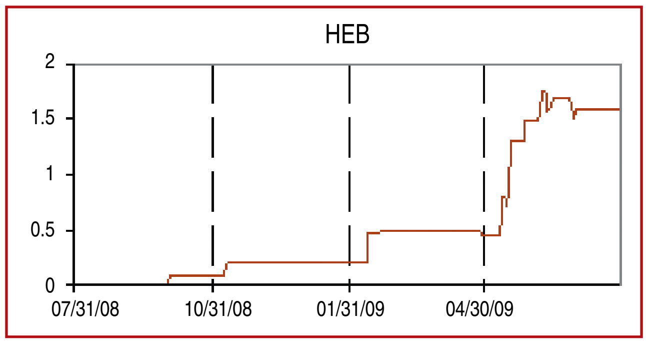


11B

FIGURE 11A and B: DATA AND FORTUNE FOR PWR ENDING JUNE 10, 2009. The least efficient is PWR, optimal T=1.80, optimal C=0.12, AP=9.7, LDF=0.105, profit ratio 0.51, 206 trades and $0.51 per trade. Its optimal T and C values, AP and profit ratio values are all near average, suggesting something inherent in this data is responsible for the filter's relatively poor performance.



12A



12B

FIGURE 12A and B: DATA AND FORTUNE FOR HEB ENDING 7/29/09. The stock having the fewest trades and best dollar return is HEB at about $88 per trade. Optimal T=3.94, optimal C=2.38, AP=18.9, LDF=1.59, profit ratio 0.61, and is 18 trades and $88.28 per trade.

SUGGESTED READING

Martinelli, Rick, and Neil Rhoads [2010]. "Predicting Market Data Using The Kalman Filter," part 1, *Technical Analysis of* STOCKS & COMMODITIES, Volume 28: January.

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

_____ [2008]. "Linear Estimation And The Kalman Filter," www.haikulabs.com/LinEst12.htm.

S&C