Trading Model Uncertainty and Statistical Process Control

John F.O. Bilson
Andrew Kumiega
Ben Van Vliet

Abstract

This paper examines the use of statistical process control (SPC) as a methodology for monitoring trading model uncertainty. Traditional quantitative risk management methods do not incorporate the inherent process control problems financial modeling. The reference adaptations apply to the *a priori* model design process and *a posteriori* model control. To build and monitor trading models, SPC can be used in conjunction with classical financial metrics to better control market risk.

JEL Classification: G11.

Key Words

Model uncertainty, statistical process control, automated trading, backtesting.

Andrew Lo and Mark Mueller [2010] have outlined new categories of uncertainty. There framework describes a perspective of economic behavior that is an alternative to the mathematical precision demanded by the natural sciences, what they call “physics envy.” Their categories are:

Level 1: **Complete Certainty.** Outcomes are deterministic. Uncertainty does not exist.

Level 2: **Risk without Uncertainty.** Known outcomes come from a known probability distribution. Through *a priori* deduction, no empirical data and no statistical inference are necessary. The probabilities have been derived.

Level 3: **Fully Reducible Uncertainty.** In Level 3 uncertainty is due to unknown probability distributions. Given large amounts of empirical data, *a posteriori* statistical inference can arrive at outcomes and probabilities close to those derived in Level 2.

Level 4: **Partially Reducible Uncertainty.** At this greater level of uncertainty, data generation processes are ambiguous. While significant empirical inference may explain some of the outcomes and their probabilities *ab ambiguitate*, a significant amount of unexplainable uncertainty exists. This leads to model uncertainty.

Level 5: **Irreducible Uncertainty.** At this highest level of uncertainty (just short of total abandon and despair) there exists a state of total ignorance. No amount of deduction or inference can overcome a state which is beyond reasoning.

By disambiguating these levels of uncertainty, we can address methods of justification appropriate for each. For quantitative finance, the relevant question Lo and Mueller ask aims at first principles: on what basis does one justify one's belief in the repeatability of a model's outputs? The problem is not unique to finance.

Other industries produce outputs which contain uncertainty. Manufacturing processes churn out physical parts that contain uncertain variations in size, but which are nevertheless within tolerance. The uncertainty in the size of machined parts is akin to Level 3. Manufacturing engineers use statistical controls to indicate when the machine is producing parts that no longer meet specification, that is, when the uncertainty has moved too far in the direction of Level 4. Unlike the clear boundaries between other levels, Level 3 is a continuum between Levels 2 and 4. Similar to machines, most financial models run in Level 3 due to uncontrollable environmental factors. The inputs and outputs of a given trading algorithm can at times be closer to Level 2 and at other times be closer to Level 4.

The closer the model outputs are to Level 2 the more reliable its outputs; the closer to Level 4, the less reliable its outputs. Statistical controls can indicate when an algorithm is no longer within its Level 3 tolerance, or specification, limits. At some critical level, when performance exceeds a specification limit, belief in the repeatability of the outputs can no longer be justified. Then, the model must be fixed, or a new model employed.

Trading is moving, and in large part already has moved, from manual control to computer numerical control (CNC). On every major securities and derivatives exchange, over 70% the trading volume is generated by computerized trade execution or black-box trading systems (Sussman et al. [2009]). The statistical theory behind the development of trading models is heavily reliant on the assumption that the underlying process is stationary. However, within the Lo and Mueller framework, these models operate in the fully reducible level of uncertainty (Level 3). Further complicating the engineering of these models, market microstructure can become ambiguous (see Easley and O'Hara [2010], and Epstein and Schneider [2008]). Trading model behavior can become uncertain. An algorithm that was successful in the past may abruptly fail. In this paper, we demonstrate that statistical process control tools used in manufacturing can be successfully adapted to identify shifts in market microstructure that affect trading models.

The traditional risk metrics—Sharpe [1994] and Sortino [1994] ratios, performance attribution analysis (see Brinson [1985]), and Value at Risk (Jorion [2001]), and draw-down analysis (see Harding et al. [2003])—lack the ability to notify risk managers that the market conditions have shifted the performance of the algorithm from Level 2 to Level 4. Once in Level 4, the stationary risk measures no longer apply. Although trading model outputs can be monitored with process control techniques, the literature does not define a standardized framework for assessing financial algorithm performance relative to design specifications.[1]

We believe the lack of process control in the new age of systematic finance is a result of the antiquated view that it is the operator (i.e. trader or portfolio manager) who is adding value. But, the trader cannot add value when trades take place in milliseconds. Another incorrect perspective is that a model can always evolve at a speed sufficient to continuously generate alpha. This has led to research that focuses on the creation of faster, more adaptive algorithms. However, recent research is identifying periods of ambiguity in the markets that result in model uncertainty (see Epstein and Schneider

[1] As with all machines, diagnostic performance measurement tools ought to be designed into the internal trade selection and position management algorithms of trading models. Ongoing performance should then be consistent with specification limits (see Juran [1951]).

[2008], and Hansen and Sargent [2006]). New risk management tools for trading models are needed to properly control algorithms during periods of ambiguity and model uncertainty.

In this paper we discuss two types of Level 3 trading model risks by example. The first type is the risk of creating a trading model that effectively over-fits the data. The results falsely show the outputs closer to Level 2, when in fact they are closer to Level 4. The example we chose is a simple pairs trading algorithm. In practice, simple pairs trading algorithms often work in-sample and do not work out-of-sample when evaluated using standard risk measures. We understand that more advanced pairs trading systems can produce excess returns, but the purpose of the example is to use SPC to identify a broken algorithm. Therefore, we were forced to use a model that does not work.

The second type is the risk of running an trading model after the model drifts toward Level 4 and stops working according to specification. In this case, we selected an example algorithm that worked over a long period of time. Then it stopped working due to a market structure shift. Statistical process control properly identified the shift in the trading model's outputs. In both examples, SPC properly identified failures that were not identified with standard backtesting and risk methodologies.

I. Model

Shewhart [1931] and Deming [1982] designed statistical methods to control machines with uncertain outputs. We apply their techniques to the control of uncertain trading models. Trading model performance can be broken down into common cause and special cause variation, since at best a model can explain 70% of the variation in financial data (see Roll [1988]). Thus, the absence of special cause variation means that a trading model's performance (e.g. its ability to produce excess returns) will be somewhere in the Level 3 continuum, preferably closer to Level 2 than Level 4. Should, however, model performance move outside of acceptable limits (that is, toward Level 4) control mechanisms will indicate deteriorating unreliability.

In the first example, working at the highest level of control, we transform a standard period P&L series into average returns.[2] These sample period returns will be graphed on what is called an X-bar chart. Suppose we collect, as an example, 5-day observations on the total P&L.

$$P\,\&\,L_t = \frac{1}{n}\sum_{i=1}^{n}\Pi_i \tag{1}$$

We let the return on an trading model be the time series of period-ending ending account values, V_t, as per:

$$V_t = V_{t-1} + P\,\&\,L_t \tag{2}$$

The rate of return on the account for time period t is defined as x_t:

$$x_t = \ln\left(\frac{V_t}{V_{t-1}}\right) \tag{3}$$

Next, suppose we calculate the n-period (e.g. 5-second, or day) sample average rates of return $\bar{x}$.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

---

[2] For clarity we selected a five day sample for both examples discussed. In practice the sample period is chosen relative the average holding period. For high-frequency systems,second-to-second, or minute-to-minute periods are generally chosen. Using samples is a key element of SPC, though not often used in quantitative finance. In finance we often assume that the asset return process is normal without sampling. With trading models, we perform 100% inspection of trade data, but nevertheless perform sampling to force the outputs closer to normality through the central limit theorem (see Shewhart [1939]).

From these values, we calculate the average return of k (e.g. 52-weeks) samples, $\bar{\bar{x}}$.

$$\bar{\bar{x}} = \frac{1}{k}\sum_{j=1}^{k}\bar{x}_j \qquad (5)$$

Through the central limit theorem, these subgroup averages, $\bar{x}_j$, will tend to be normally distributed around $\bar{\bar{x}}$ with dispersion $s_{\bar{x}}$. In practice, because n < 10, we use the sample range to estimate $s_{\bar{x}}$. Range{ $x_1,\ldots,x_k$ }, the range R of returns in a sample is defined as:

$$R = x_{(n)} - x_{(1)} \qquad (6)$$

And, the average range $\bar{R}$ over k samples is:

$$\bar{R} = \frac{1}{k}\sum_{j=1}^{k} R_j \qquad (7)$$

The relationship between the average range and the standard deviation depends only on the sample size n (see Patnaik 1950). The process standard deviation s_x and the standard deviation of sample means $s_{\bar{x}}$ are estimated using d_2, a constant based on the subgroup size per Exhibit 1, as follows:

$$s_x = \frac{\bar{R}}{d_2} \qquad \text{and,} \qquad s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \qquad (8)$$

From an absolute return perspective, the annualized expected return $E(r)$ is equal to the sample average of backtested or historic returns of the trading system $\bar{\bar{x}}$ times the number of periods in a year $n \cdot k$, say 260 trading days[3]. The annualized standard deviation of sample returns s_r is the sample standard deviation times the square root of $n \cdot k$.

[3] Including holidays, the actual number of trading days per year is around 252. However, to simplify we will assume 5 trading days per week times 52 weeks is 260 trading days per year.

This enables us to define performance boundaries. As long as $\bar{x}$ performance remains within these boundaries (i.e. as long as the hypothesis cannot be rejected), we are justified in sustaining our belief in the repeatability (i.e. the future performance) of the trading system. An upper control limit (UCL) and lower control limit (LCL) define conditions for when belief that future performance will be consistent with past performance can no longer be sustained. These control limits are three standard deviations above and below the expected return, as estimated by:

$$\begin{aligned} UCL_{\bar{x}} &= \bar{\bar{x}} + A_2 \cdot \bar{R} \quad \approx \bar{\bar{x}} + 3 \cdot s_{\bar{x}} \\ LCL_{\bar{x}} &= \bar{\bar{x}} - A_2 \cdot \bar{R} \quad \approx \bar{\bar{x}} - 3 \cdot s_{\bar{x}} \end{aligned} \tag{9}$$

Where A_2 is the anti-biasing constant per Exhibit 1, and:

$$A_2 = \frac{3}{d_2 \sqrt{n}} \tag{10}$$

Value of *c*	**Basic Factors**		**Factors for Averages**			**Factors for Ranges**			
	d_2	d_3	A	A_1	A_2	D_1	D_2	D_3	D_4
2	1.128	0.853	2.121	3.760	1.880	0	3.686	0	3.267
3	1.693	0.888	1.732	2.394	1.023	0	4.358	0	2.575
4	2.059	0.880	1.500	1.880	0.729	0	4.698	0	2.282
5	2.326	0.864	1.342	1.596	0.557	0	4.918	0	2.115

Exhibit 1: Statistical Constants

To assess the risk in a trading model, we use sample ranges. These sample period ranges will be graphed on what is called an R chart. The true standard deviation of returns is found from the distribution of $w = R / \sigma$. The standard deviation of w is d_3 and is a function of the sample size n. Since $R = w \cdot \sigma$, and because the true standard deviation, we can estimate s_R by

$$s_R = d_3 \cdot \frac{\bar{R}}{d_2} \tag{11}$$

As a result, the control limits for R are three standard deviations above and below the expected return, as estimated by:

$$
\begin{aligned}
UCL_R &= D_4 \cdot \overline{R} \\
LCL_R &= D_3 \cdot \overline{R}
\end{aligned}
\tag{12}
$$

The estimation of s_x in (10) also allows us to represent the annualized standard deviation as:

$$
\sigma_x = \frac{\overline{R}}{d_2} \cdot \sqrt{n \cdot k} \tag{13}
$$

Again, we now have risk performance boundaries. As long as the sample ranges, R, performance remains within these boundaries; we are justified in sustaining our belief in the future risk of the trading model. The upper and lower control limits (UCL and LCL) define conditions for when belief that future risk will be consistent with past risk can no longer be sustained.

II. Automated Trading Systems

For both the risk and return of a trading model, the mean and control limit values (UCL and LCL) for these metrics is calculated using sampled performance data from the backtest. Data points falling outside of the control limits represent a shift in the process. Nelson [1984] defines eight signals in total:

- Any single measurement above or below the 3 standard deviation UCL or LCL.
- 9 points in a row on one side of the mean.
- 6 points in a row increasing or decreasing.
- 14 points in a row toggling back and forth between increasing and decreasing.
- 2 out of 3 points in a row more than 2 standard deviations from the mean in the same direction.
- 4 out of 5 points in a row more than 1 standard deviation from the mean in the same direction.
- 15 points in a row all within plus or minus 1 standard deviation.
- 8 points in a row all outside 1 standard deviation in either direction.

Proper backtesting should confirm the validity and accuracy of a trading system's algorithms, and the repeatability of its output process capabilities. If a trading model is in control, this does not mean the absence of variation in the process, only that model outputs are within specified tolerance levels (see Pyzdek [2003]). In quality control, every process has variation. The goal is to identify the attributes that vary, and then to minimize their variation. From the quality control perspective, even algorithms in Level 1 have variation in their performance. Such variation could be, for example, the number of trades per hour. Also, variation in technological performance could lead to probabilistic outcomes of otherwise deterministic mathematics.

Trading models come in all shapes and sizes and a specific application of SPC depends on the nature of the strategy and the backtesting method used to prove the performance. Each of the metrics could be analyzed using SPC.

- Mean/Median Profit and Loss
- Software bug
- Average and Standard Deviation of Returns
- Sharpe [1994] and Sortino [1994] Ratios
- Percentage of Winning Days
- Number of Winning and Losing Trades, Winning and Losing Days
- Drawdowns (Magdon-Ismail 2003)
- Information Coefficient / Spearman Correlation [1904]
- Over/under-weights an equity sector
- Technological latency measures

By testing for control during the empirical validation step (i.e. backtest) we calculate the trading model's control limits. That is, if the system is not in control, if special variation exists in the backtest results, we cannot be justified in believing the performance results will be repeatable in a real trading environment. Furthermore, given a successful backtest with defined control limits, the occurrence of out of control performance indicates a failure of the trading model.

Once an algorithm is deemed out-of-control, the goal is to uncover the root cause of special variation in the trading model. Special cause variation is assignable to a root cause and can be eliminated through corrective action (Hossain et al. [1996]). The root causes of special, or assignable, variation in trading models may come from one of the following factors that may impact market microstructure (see Kumiega and Van Vliet [2008]):

- Legal changes, such as Sarbanes Oxley.
- Software bugs.
- Regulatory or GAAP changes.
- Changes in exchange matching algorithms.
- Technology changes, such as new server technology or messaging protocol version changes. (An trading model, unlike a physical machine, does not wear down; parts do not become corroded. Technology can, however, become obsolete.)
- Economic shifts.
- Political changes.
- Index composition changes.
- New, tradable contracts are offered.
- Better designed signals that reduce forecasting error.
- New machines coming online. SPC can alert firms to new competition, where another firm has rolled out a competing machine.

We consider two research hypotheses:

1.) Will relying on SPC metrics as opposed to the traditional finance metrics result in a different conclusion about a model's uncertainty and its ability to reproduce Sharpe Ratios out of sample;

2.) Will relying on SPC metrics for trading model uncertainty as opposed to traditional finance metrics be a better predictor of increased uncertainty and poor performance, resulting in the shutting down[4] of a working trading model sooner, leading

[4] In manufacturing, an out-of-control machine is shut off. In finance, a risk manager may reduce the capital allocated to an out-of-control trading machine, choose to accept the risk, while remaining cognizant of the necessity to close positions quickly, or permit only trades that close existing open positions. The

to a smaller loss.

With respect to question one, we find that SPC metrics provide foresight regarding trading model uncertainty and generate an investment signal contrary to that generated by the traditional methods of finance. With respect to question two, we find that SPC can trigger risk signals when a trading model drifts out of specification and toward Level 4 prior to periods of increased uncertainty and poor performance. Thus, in the two example trading systems we present, we define conditions that justify initial investment in a trading model via the backtest, but also conditions under which justification for continued investment in a working trading model can no longer be sustained, via the control limits.

II. Long-Short Trading Strategy

To illustrate the application of SPC to backtesting, our first case study is a simplified pairs trading system using statistical arbitrage. In the trading industry, statistical arbitrage of pairs (see Hassan, et al. [2010], and Van Vliet [2007]) of stocks within a particular industry group is popular because of their convergence characteristics. We selected this algorithm since simple pairs algorithms regularly perform well in a backtest and poorly in real trading. We have not attempted to create a more robust algorithm since the purpose of this example is to apply SPC to an algorithm that is working using classical finance measures and broken using SPC.

Statistical arbitrage is based on fundamental analysis, and so we selected pairs of fundamentally similar companies that could be expected to exhibit similar price behavior.

---

implication is that a good trader, like a good pilot, can land the plane safely after the autopilot is switched off. A human trader may add value during financial crises and abnormal trading environments.

To create a more stable value for the mean ratio of prices, we used a 30 day moving average, comparing the observed value of the spread between the market prices to a more stable mean value. To normalize the difference between the observed price and the 30 day moving average, our system uses a simple z-transform method:

$$\Delta_{norm} = \frac{\Delta - MA_{30}(\Delta)}{\sigma_{30}(\Delta)} \tag{14}$$

Where Δ is the observed price of the spread, the normalized prices, Δ_{norm} expresses the normally distributed price in terms of standard deviation from the mean, and the mean price $MA_{30}(\Delta)$ is the 30 day spread-price moving average.

We define a long position in the pair as short Stock A and long Stock B. A short position is long Stock A and short Stock B. When Δ_{norm} exceeds 2, we consider this a bearish signal and we expect a reversion of the spread price down to the mean of 0. As a result we take a hypothetical short position in the spread. Alternatively, when Δ_{norm} is less than –2, this is a bullish signal and we expect to see a reversion of the spread price up the mean. In this case we take a long position in the spread. We add to positions daily and hold each position for five trading days.

For simplicity, we assume that all opening trades are executed at the closing price on the day the signal is received and closing trades executed at the closing price 5 days hence. Dollar amounts assume that long positions are paid for in full and a margin of 50% is assessed on short trades. Commissions have been omitted since the purpose of this research is to show the application of SPC to a trading model, not to prove the profitability of the trading algorithm.

The selection process for the pairs consisted of a comparison of three fundamental factors: industry sector; market capitalization; and, price-to-earnings ratio. Exhibit 2 contains the list of the pairs of stocks used in our system and their sectors.

STOCK PAIRS		
Sector	**Stock A**	**Stock B**
Home Improvement Stores	Home Depot	Lowe’s
Drug Stores	Walgreen’s	CVS
Discount Variety Stores	Costco	Target
Processed & Packaged	Pepsi	Coca Cola
Major Integrated Oil & Gas	Exxon Mobil	British Petroleum
Money Center Banks	Bank of America	Citigroup
Investment Brokerage	Goldman Sachs	Merrill Lynch
Health Car Plans	United Health	Wellpoint
Drug Manufacturers	Abbott Labs	Merck
Air Delivery & Freight	UPS	FedEx
Chemicals	Dow Chemical	DuPont
Communications Equipment	Motorola	Nokia

Exhibit 2: Stock Pairs

First, we evaluated this trading machine using the traditional risk tools from finance. We tested this system using four years of daily price data on the stocks. (Daily adjusted closing price data from 1/1/2002 to 4/1/2006 was provided by Yahoo! Finance.) Over the first two years, the in sample test produced the following performance metrics:

Performance Metric	Trading System	S&P 500 Index	Nasdaq 100 Index	Long/Short Benchmark
Average Annual Return	51.57%	2.29%	1.61%	8.08%
Average Annual Volatility	23.63%	21.55%	32.45%	6.44%
Average Sharpe Ratio	2.180	.106	.049	1.254

Exhibit 3: In-sample Test Results

The values in Exhibit 3 are effectively static since they were calculated over a fixed time interval. Based upon a static Sharpe ratio, these outputs are assumed to have a stationary mean and a stable process.

Based upon the traditional performance metrics shown in Exhibit 3, the trading system produced acceptable returns. After all, these are impressive results. Based upon the backtest, this trading strategy would have been a preferable investment system compared to the other investments (e.g. the S&P 500 Index, the Nasdaq Composite Index, and the SGI WISE Long/Short Index) for the in-sample test period.

The Sharpe ratio is regularly used to indicate if a process is repeatable. The key assumption of the Sharpe ratio is that the output process is in control since the system's return to unit of risk is acceptable to the investor. Based on the Sharpe ratio of 2.18 in Exhibit 3, the process would be deemed to be in control based on this measure.

Next, we measured the outputs of this trading strategy using SPC. Our measurements are based on collecting samples of 5 days of returns to create an X-bar chart. We are concerned with whether the return process will confirm the performance results shown in Exhibit 3. If so, it will confirm that this trading system is in fact a sufficiently close to Level 2 to be considered a repeatable process. Exhibit 4 shows the X-bar chart of returns over the 2 year in-sample backtest.

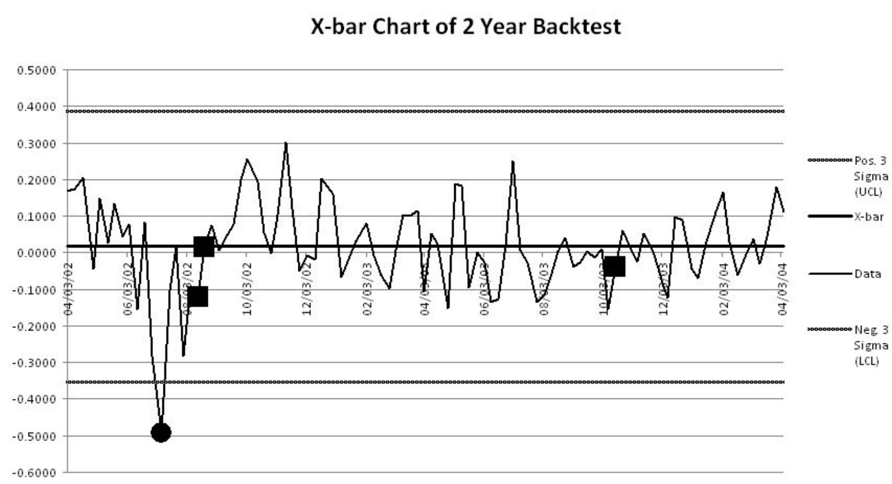

Exhibit 4: 2 Year X-bar Chart of Returns
(In Exhibit 4, the circled data point violates the 3 sigma limit. The three squares identify violations of 7 points on one side of the mean.)

While the returns are impressive, the X-bar chart shows that the returns are not stable. The trading model is therefore out-of-control. The uncertainty is close enough to Level 4 that we cannot expect the results to be repeatable out of sample. As can been seen in Exhibit 4, the chart of returns exceeds the 3 sigma LCL once, and three times has seven consecutive measurements on one side of the mean.

Out-of-sample results in Exhibit 5 show that the system posted significantly poorer returns and exhibited much greater volatility, inconsistent with the mean and standard deviation of returns experienced during the in-sample backtest. This is a strong argument for using process control metrics to analyze trading models. Thus, with respect to hypothesis one, we find that SPC more correctly indicated the nature and repeatability of the system.

Performance Metric	Trading System	S&P 500 Index	Nasdaq 100 Index	Long/Short Benchmark
Average Annual Return	5.79%	8.03%	8.14%	15.10%
Average Annual Volatility	16.86%	10.49%	15.09%	5.60%
Average Sharpe Ratio	.34	.766	.539	2.697

Exhibit 5: Out-of-Sample Test Results

III. Foreign Currency Trading Strategy

To illustrate the use of SPC control charts to evaluate a working trading model, we use a well-known foreign currency carry strategy (see Bilson and Hseih [1987], and Fama [1984]). This example was selected since it performed well both in-sample and out of sample. The algrotihm then stopped working during the market turmoil of 2008-2009. First, we define S_t to be the exchange rate expressed as the USD price of a unit of foreign currency. Second, i_t* is the foreign interest rate and i_t is the USD interest rate. Now suppose we borrow 1 USD, convert it into the foreign currency, earn the foreign interest rate, and then convert it back to USD. The profit, r_t, on this trade is:

$$r_t = \frac{S_t}{S_{t-1}}(1+i^*_{t-1}) - (1+i_{t-1}) \tag{15}$$

This can be written as:

$$r_t = \frac{(S_t - S_{t-1})}{S_{t-1}}(1+i^*_{t-1}) + (i^*_{t-1} - i_{t-1}) \tag{16}$$

This says that the excess return can be written as the capital gain or loss on the end of period foreign currency position and the interest rate differential. The basic idea behind this carry trade is that the first term is unpredictable while the second term is known at the time of the investment. Assuming that the expected capital gain or loss is zero, we obtain:

$$E(r_t) = (i^*_{t-1} - i_{t-1}) \tag{17}$$

Define *x(n,t)* as the interest rate differential between currency *n* and the USD. In each period, we calculate the average value of *x(n,t)* and its standard deviation. We then compute the standardized interest rate differential:

$$z_{n,t} = \frac{[x_{n,t} - \bar{x}_t]}{\sigma_t} \tag{18}$$

The position taken in currency *n* in period *t* is given by the logistic function:

$$q_{n,t} = \frac{\exp(z_{n,t}) - 1}{1 + \exp(z_{n,t})} \tag{19}$$

While $z_{n,t}$ is a standardized variable that will range between, say, -5 and +5, qn,t will range between -100% and +100%. So we can think of $q_{n,t}$ is the proportion of capital invested in currency *n* at time *t*. An investor with a higher degree of risk tolerance could multiply these positions by a number greater than unity to obtain higher returns and higher risk. The following table provides a position sheet for 10/31/2008.

Position Sheet: 10/31/2008					
Currency	**Short**	**Yield**	**Currency**	**Long**	**Yield**
XAU	-56%	0.00%	ZAR	77%	12.20%
XAG	-52%	0.50%	HUF	64%	10.25%
JPY	-50%	0.70%	MXN	40%	7.75%
CHF	-38%	1.75%	NZD	23%	6.40%
CAD	-30%	2.40%	AUD	20%	6.14%
CZK	-16%	3.50%	USD	19%	1.00%
EUR	-12%	3.80%	PLN	13%	5.63%
GBP	-2%	4.50%			
TOTAL	**-256%**	**3.10%**	**TOTAL**	**256%**	**22.68%**

Exhibit 6: Currency Trading Positions

The portfolio is borrowing funds in Gold, Silver, Yen, Swiss, Canada, Czech, Euros and Sterling. Exhibit 6 shows that the total borrowings amount to 256% of capital and the cost of funds is 3.10%. The program is investing in South Africa, Hungary, Mexico, New Zealand, Australia, the U.S. and Poland. The average yield on investments is 22.68%. If none of the exchange rates changed, the excess return on the portfolio would be: 22.68% — 3.10% = 19.57% per annum.

There is, of course, no economic reason why the carry trade should work (Bilson [1981]). The strategy basically involves borrowing in rich, stable currencies and lending in less developed, high yielding currencies. If a financial crisis should occur, the carry trade could break down (see Brunnermeier et al. [2008]).

Unlike the statistical arbitrage trading system, this system remained in control over the duration of the backtest period, which permitted the allocation of investment capital. Over the course of trading, the system made positive profits with variation within the specified control limits.

We examined both X-bar and R charts for statistical control, where the range is a proxy for the variation (or risk) of the system. We selected R charts since, again as

Brunnermeier et al. [2008] note, large ranges can occur in currency carry trades and result in massive profit and loss swings from the panic unwinding of trades.

As can be seen in Exhibit 7 and 8, after a long period, the system did eventually break down.

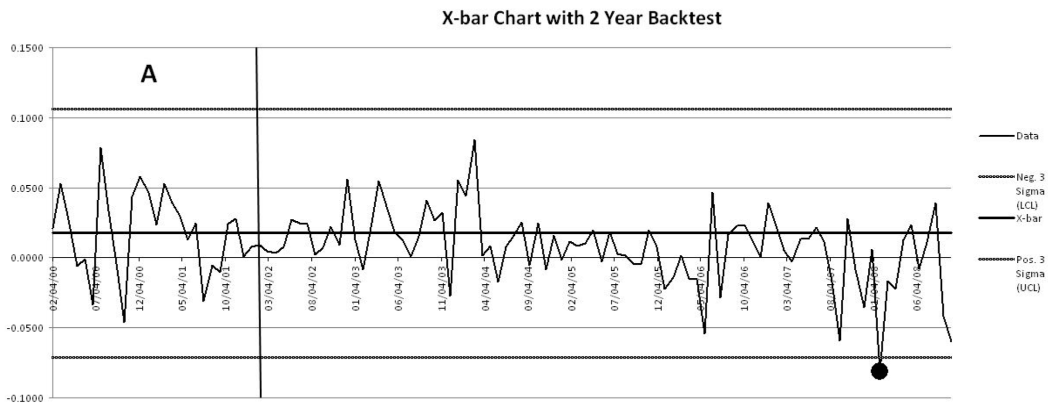

Exhibit 7: X-bar Chart of Returns with 2 Year Backtest
(In Exhibit 7, the circled point violates the 3 sigma limit. Period A is the backtest period.)

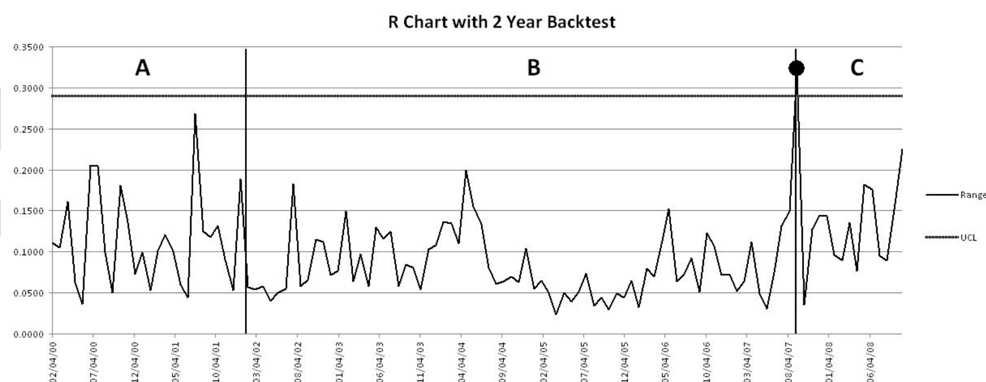

Exhibit 8 R Chart of Returns with 2 Year Backtest
(In Exhibit 8, the circled point violates the 3 sigma limit. Period A is the backtest period; Period B is the out of sample, in control period; and, Period C is the out of control period)

The X-bar chart remains in control until 01/25/08. The R chart shows an out of control at 09/07/07. In both cases, however, the out of control signal portends the steep decline in account value as can be seen in Exhibit 9.

The earlier R chart out of control signal would have shut down the system on 09/07/07 and Exhibit 9 shows the value of a $100 investment over the three periods—(A) the two year backtest, (B) the out of sample period where the trading model remains within statistical control, and (C) the period after the R chart triggered an out of control signal. After the R chart out of control signal, the system lost 48% from that time until the end of the data.

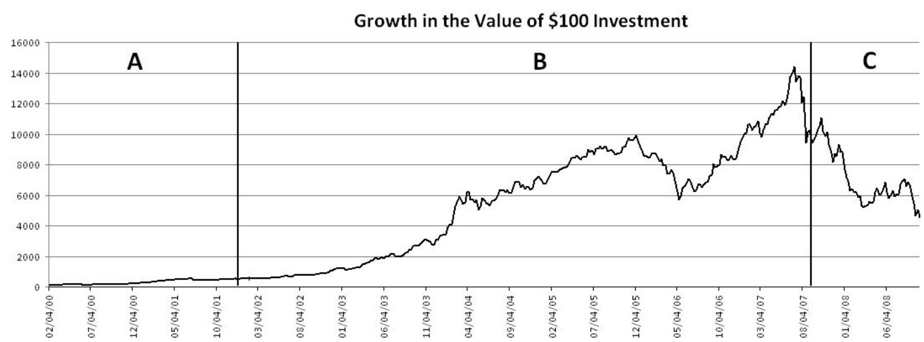

Exhibit 9: Growth in the Value of $100
(In Exhibit 9, Period A is the backtest period; Period B is the out of sample, in control period; and, Period C is the out of control period.)

Exhibit 10 shows the returns, standard deviations and Sharpe ratios for the three periods (A, B, and C) shown in Exhibit 9.

Performance Metric	**2 Year Backtest (Period A)**	**Out of Sample, In Control Period (Period B)**	**Out of Control Period (Period C)**
Annual Return	93.73%	55.28%	-48.41%
Annual Volatility	41.27%	31.78%	44.56%
Sharpe Ratio	2.27	1.74	-1.09

Exhibit 10: Performance Results

IV. Conclusion

Researchers such as Lo and Mueller [2010], Easley and O’Hara [2010], Epstein and Schneider [2008], and prior to them Chow and Sarin [2002], have described risk in new terms of uncertainty and ambiguity. Periods of uncertainty drive financial models to perform outside of design specifications. We have shown through two examples how

SPC can be used to select and control trading models during uncertain times. We find that SPC generates meaningful investment signals contrary to the signals generated by traditional finance metrics. This confirms both research hypotheses one and two.

The conclusions of this research contradict standard operating procedures of financial engineering which seeks to maximize the P&L of a given trading model across many products and all market conditions. This standard requires that trading models quickly adjust to the changing market environment. The research method presented allows financial engineers to properly monitor the models' dynamic outputs to determine if the model is working to specifications.

The application of manufacturing process control to financial algorithms should hasten the shift in finance from algorithms that are optimized in the standard procedure to algorithms that are optimized for specific environments. This shift is similar to that in manufacturing in the 1970s, from integrated production lines to lean manufacturing. In lean manufacturing, machining is performed in small cells that are optimized for a specific family of parts which are monitored using real-time SPC. Statistical process control enables the shift to the creation of portfolio of trading models, where each model can turned on or off according to its ability to perform to specification. Thus, in-control algorithms are allocated trading capital.

The key difference between manufacturing and algorithmic trading is that in trading, algorithms can be left to run continuously, yet limiting their ability to execute trades (i.e. paper trading). This allows an engineer to determine if the algorithm is working in the current market environment. This also allows an engineer to benchmark competing models in real time. Further, drawdowns from shifts in market microstructures will decrease through timely detection with SPC. We believe that with the addition of SPC to algorithm development the risk to reward ratio for the entire portfolio of algorithms will increase. We expect the way algorithms are designed will shift toward what is the current manufacturing standard of specialized manufacturing cells for specific parts monitored using real-time SPC, versus manually controlled general machines over a broad range of products.