

# **Developing Robust Trading Systems, with Implications for Position Sizing and System Health**

Howard B. Bandy, Ph.D.

Submission to the NAAIM Wagner Paper Award, February 2012

## **The Goal**

As traders and investors, particularly as active investors, it is important to have a high level of confidence that:

- The trading system being used is robust.
- The trading system is healthy.
- Trades are being taken in the correct size to produce the fastest equity growth while keeping drawdown below an acceptable percentage.
- There is a plan for dealing with drawdowns.

This paper describes unique and practical techniques for gaining that confidence, and illustrates with a fully analyzed trading system.

## **Trading Systems – Some Background**

Any purchase made or position taken can be divided into one of three categories:

- An investment. If you expect to hold a position forever and have your lawyer add it to the list of assets in your will, it is an investment. Perhaps a dream vacation property that will be used for generations to come.
- An expense. If you expect it to be consumed or wear out and have little or no residual value, it is an expense. A leased office suite, for example.
- A trade. If you expect to dispose of it, hopefully at a profit, within your lifetime, it is a trade. Certainly most stock positions, even if the planned holding period is years, and probably most real estate.

The process by which issues to buy and sell are selected, and the timing of those actions is determined, is a trading system. This working definition of a trading system is probably broad enough to apply to almost all methods used by readers of this paper.

Focusing on securities, each trade begins with an entry into the position with expectation that the position will be more valuable in the future.

There is a reason each position is entered. Whether that reason is based on personal experience, fundamental analysis, chart analysis, expert opinion, a set of logical rules, a computed indicator, or for any other reason, there is recognition of some condition or pattern that precedes profitable opportunities.

Eventually the position is closed out. There are five general reasons to exit from a position:

- Recognition of some condition or pattern indicating that additional profit is unlikely.
- A profit level, such that additional profit is unlikely, has been reached.
- A maximum holding period, such that holding longer is unlikely to bring additional profit, has been reached.
- A trailing stop – such as the chandelier stop, or the parabolic component of the SAR system.
- A maximum loss stop, with a "last resort" exit price.

Each trade consists of:

1. Recognizing the pattern in the data that signals the entry.
2. Taking the position.
3. Exiting the trade based on one of the five criteria.

Every trading system can be defined as a combination of a model and a data stream.

- The model contains the rules – the logic—that define the patterns. Models are relatively rigid. While there may be some capability for the model to adapt to changes in the data, there are limits.

Each model is designed to recognize a specific pattern, such as:

- Prices are trending.
- Prices are mean-reverting.
- Seasonality is favorable.
- A trusted advisor has issued a report.
- The data is the price and volume of the issue being traded, together with auxiliary data used by the model. The data consists of a combination of signal and noise. The signal component is the pattern sequence the model is designed to recognize. The noise component is everything not specifically included in the signal – even if it contains profitable patterns that could be detected by some other model.

## **Synchronization**

A trading system is profitable only when the model and the data are synchronized – when the patterns identified by the logic actually do precede profitable trading opportunities.

The process of trading system development consists of several phases:

1. Conceiving an idea and defining the rules.
2. Defining the criteria by which system performance will be measured. This is the objective function or fitness function. It incorporates those characteristics that are important to the trader, such as profitability, drawdown, holding period, trading frequency.
3. Using historical data to adjust the model so that it is synchronized to the data. This is the in-sample (IS) backtesting and optimization phase.

4. Using a different, usually more recent, set of historical data to verify that the model is identifying key signals that precede profitable trades and not just fitting the model to the specific test data. This is the out-of-sample (OOS) validation phase, often including walk forward tests. The results of validation tests are a measure of the synchronization between the model and the data.
5. Provided the out-of-sample performance is satisfactory, moving the system from development to trading. At the time development is complete, the model and the data are synchronized.
6. Determining appropriate position size.
7. Monitoring the health of the system during live trading, adjusting position size as necessary.

### **Walk Forward Testing**

The walk forward process is one of the best ways to learn how the system will perform in the future.

The procedure begins with selection of an objective function by which trading system performance is evaluated. The purpose of the objective function is to incorporate the subjective criteria of the trader, quantify that into a single-valued metric, and enable alternative systems to be objectively ranked and compared.

Several, perhaps many, test runs are made to determine the settings of the logic and parameters of the model, and to determine the length of time the data remains relatively consistent. Knowledge of that length is important because the model must synchronize itself to the data during the earlier portion, then signal trades during the later portion. The length of time it takes for the system to synchronize determines the length of the in-sample period. The additional length of time that the system continues to give accurate signals determines the length of the out-of-sample period. The system must be resynchronized periodically, with the length of the out-of-sample period corresponding to the maximum time between resynchronizations.

Knowing the length of the in-sample period and the length of the out-of-sample period, and having the objective function to rank alternatives, the walk forward process consists of a number of steps.

1. Initialize the process by setting the in-sample period length, the out-of-sample period length, and the dates of the first step:
  - a. Set the date for the beginning of the first in-sample period.
  - b. Set the length of the in-sample period (the length of subsequent periods will all be the same), which will determine the date of the end of the first in-sample period.
  - c. The out-of-sample period begins immediately following the in-sample period. (Using data earlier than the in-sample period as an out-of-sample period is poor practice. It results in overestimates of profit and underestimates of risk.)

2. Repeat these steps until all of the data has been processed:
  - a. Search through many alternative sets of rules and parameter values using the in-sample data. This is typically done using the optimization features of the development platform.
  - b. Evaluate the performance of each alternative and give it a score using the objective function.
  - c. Sort the alternatives by objective function score.
  - d. Using the rules and parameters of the highest ranked alternative, apply the trading system to the out-of-sample data. This is a one time run. Store the results for later use. Since the process is automatic, and you have no way to examine the performance of any of the alternatives other than the top-ranked one, you must have confidence that your objective function accurately reflects your criteria.
  - e. Move the in-sample and out-of-sample periods forward by the length of the out-of-sample period.
3. Gather together, then analyze, the results from all of the out-of-sample tests.
  - a. If the performance is satisfactory, then the system has passed the walk forward test and is ready to be traded. Use the parameter values chosen in the final in-sample period.
  - b. If the performance is not satisfactory, return to earlier development stages and revise the rules and parameters.

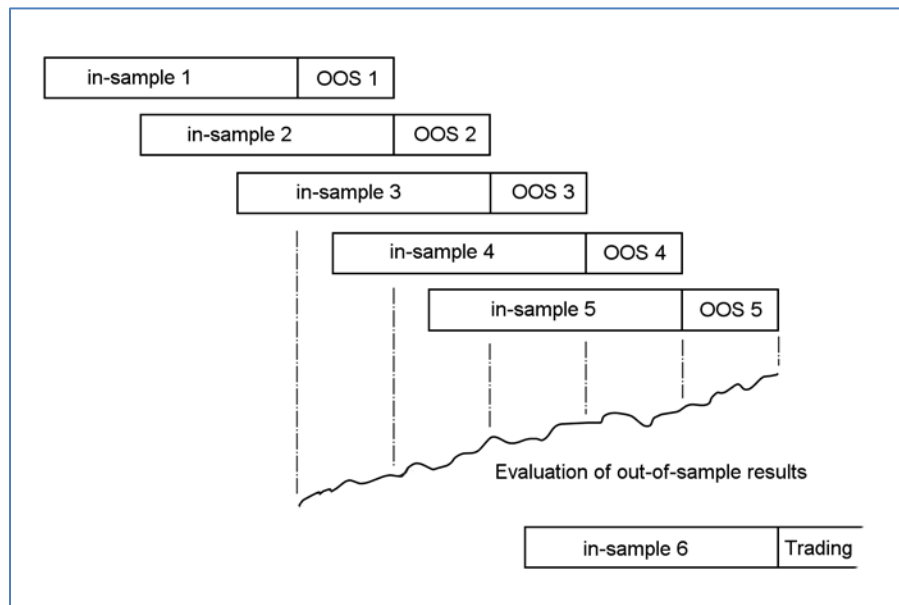


Figure 1 – The walk forward process

The walk forward process:

- Gives "practice" opportunities to observe the action of the system as it moves from development to trading – in-sample optimization to live trading using the best parameter values.
- Provides the best data available to estimate future performance of the system.

The walk forward process shows how the system reacts to, and adapts to, changes in the market being traded. Study of the periods when out-of-sample results are good and when they are poor can provide information useful as conditions change during live trading.

It is important that the logic both have the ability to adapt, and be allowed to adapt. It is rare that the system stays in synchronization over a long period. Models with fixed parameters are forced to compromise, fitting one period well and another not at all, or fitting several periods poorly.

### **Best Estimate Trade Set**

Estimates of future performance, assessment of system health, and computation of position size are all based on a "best estimate" set of data. As the name implies, that data should be the best estimate and least biased data available. If the system is being developed for a single tradable issue, and walk forward validation runs have been made, the set of trades produced in the out-of-sample periods is probably the best available. If the system is being developed for a portfolio, and there are periods when fewer than the maximum permitted number of positions are being held, there may be serial correlation in the out-of-sample results. This often happens with mean reversion systems where trades occur in bunches as many of the issues have spikes or dips together. One way to try to remove the effect of serial correlation is to use daily changes in the equity of the trading account in place of individual trades.

The equity curve associated with the best estimate set gives a single estimate of system performance – that is, it gives a single estimate of account growth and of drawdown.

There will always be a bias in any estimate of future performance. No matter how we try to maintain strict impartiality in selecting issues to trade, rules and logic, and parameter values, developers invariably select for success. Additional bias comes from using a single value for compound rate of return (CAR) and for maximum drawdown (MaxDD).

### **Distributions**

It is unlikely that future data will be identical to the data used during system development. Patterns will become clearer and less clear, more frequent and less frequent; trades will be shorter and longer; conditions will occur in different proportions and in different order. Even though the best estimate set represents only a single sequence, Monte Carlo simulation can be used to create many alternative sequences from that same data, and those sequences can be used to estimate the distributions of CAR and MaxDD. Knowing the distributions allows the

developer to create many possible hypothetical future trade sequences, all equally likely, and from them estimate the range of CAR and of MaxDD.

## **Drawdown**

The primary reason traders stop trading systems is that the drawdown exceeds their personal tolerance for risk. By knowing the probabilities of drawdowns of various magnitudes, the trader can calculate the position size that gives maximum account growth while limiting the drawdown to an acceptable level.

## **Position Sizing**

Position sizing refers to the size of each trade. For futures that could mean the number of contracts. For stocks, mutual funds, and exchange traded funds that could mean the percentage of available equity used for each position.

Ralph Vince has written five books about money management. He has popularized the technique, called fixed fraction, of allocating a fraction of the trading account to each trade. He describes methods for computing the fraction, called optimal  $f$ , that maximizes the expected value of the account at the end of some period. (Following Vince's notation, the ratio of the final account balance to the initial balance is called terminal wealth relative, or TWR.) He discusses the relationship between trading at various fractions, account growth, and drawdown. His work is excellent and highly recommended.

Calculation of optimal  $f$  is based on the largest losing trade—a single value. If, in trading, a larger losing trade is experienced, then the fraction changes. But calculation of the fraction does not change when other important metrics of the system change – such as win to loss ratio and percentage of trades that are winners.

The technique described in this paper also determines position as a fraction of available funds. The unique feature is that it takes the distribution of trades and overall system health into account. When risk increases, position size is decreased; and when risk decreases, position size is increased.

## **System Health**

One of the questions often heard among traders is how to tell whether a system is working or is broken. All large drawdowns begin with small drawdowns. At what point should the trader worry? When is it probably safe to continue to trade through a drawdown? When should the system be taken offline and paper traded? When is it safe to resume trading? When should it be retired from service and sent back to be re-developed?

There are some techniques drawn from classical statistical analysis that can be used, in some circumstances, to evaluate system health. As with many data analysis procedures, there is a

tradeoff between the size of the sample being tested and the confidence in the result. Since trades occur in a time sequence and new data comes only with additional trades, enlarging the sample lengthens the lag between the time a system began to fail and the time the tests gave conclusive evidence that it had failed. (Use of classical statistical tests is not addressed in this paper.)

A welcome aspect of the position sizing technique I describe is that it also measures system health. When the system begins to deteriorate, the recommended position size is automatically reduced. Eventually, as the system fails completely, the recommended position size is zero. That is appropriate, since the correct position size for a system that is broken is zero. Do not trade through steep drawdowns. There is no way to determine whether a small drawdown is temporary and the system will return to synchronization soon, or if it is the beginning of a large drawdown from which it never recovers.

### Example using an Actual Trading System

In keeping with the selection criteria prescribed by the Call for Papers for the Wagner Award, this paper includes an example of a trading system that:

- Outperforms the market.
- Includes opportunity for analysis of parameter sensitivity and robustness.
- Is disclosed in complete detail so that any reader can replicate these results.
- Is supported by accepted modeling and simulation procedures.

The system:

- Is mean reverting. (It buys dips and shorts spikes.)
- Trades SPY (also most ETFs and many stocks) using end-of-day data.
- Trades both long and short.
- Is selective, taking only trades with a high probability of profit.
- Takes all of its actions at the close of trading of the regular session.

Figure 2 shows the AmiBroker code. The major sections of the program are identified, showing the system settings, code for long signals, code for short signals, and plotting statements.

```
//      NAAIMZScore.af1
//
//      Howard Bandy
//      www.blueowlpress.com
//
//      February 2012
//
//      AmiBroker code implementing the
//      trading system analyzed in the
//      paper submitted to the Wagner Competition
//
//      Buy low z-score
//      Short high z-score
//
//
//      //////////// System settings ////////////
```

```

OptimizerSetEngine( "cmae" );

SetTradeDelays( 0, 0, 0, 0 );
BuyPrice = Close;
SellPrice = Close;

MaxPos = 10;
SetOption( "MaxOpenPositions", MaxPos );
SetOption( "InitialEquity", 100000 );
SetPositionSize( 10000, spsValue );

SetBacktestMode( backtestregulararawmulti );

////////// Long code //////////

LongFilterLength = 1; //Optimize("LongFilterLength",190,10,200,10);
LongFilter = C >= EMA( C, LongFilterLength );

LongZLength = Optimize( "LongZLength", 10, 2, 20, 1 );
LongZScore = ( C - EMA( C, LongZLength ) ) / StDev( C, LongZLength );

BuyLevel = Optimize( "BuyLevel", -0.2, -5.0, 0, 0.1 );
BuySecond = Optimize( "BuySecond", 0.7, 0.1, 1.0, 0.1 );

Buy = LongFilter AND Cross( BuyLevel, LongZScore );
Buy = LongFilter AND Cross( BuyLevel - BuySecond, LongZScore );

Sell = Cross( LongZScore, BuyLevel );

////////// Short Code //////////

ShortFilterLength = 1; //Optimize("ShortFilterLength",180,10,200,10);
ShortFilter = C <= EMA( C, ShortFilterLength );

ShortZLength = Optimize( "ShortZLength", 20, 2, 20, 1 );
ShortZScore = ( C - EMA( C, ShortZLength ) ) / StDev( C, ShortZLength );

ShortLevel = Optimize( "ShortLevel", 0.1, 0.0, 5.0, 0.1 );
ShortSecond = Optimize( "ShortSecond", 0.4, 0.1, 1.0, 0.1 );

Short = ShortFilter AND Cross( ShortZScore, ShortLevel );
Short = ShortFilter AND Cross( ShortZScore, ShortLevel + ShortSecond );

Cover = Cross( ShortLevel, ShortZScore );

////////// Plot //////////

Plot( C, "C", colorBlack, styleCandle );
shapes = IIf( Buy, shapeUpArrow, IIf( Sell, shapeDownArrow, shapeNone ) );
shapecolors = IIf( Buy, colorGreen, IIf( Sell, colorRed, colorWhite ) );
PlotShapes( shapes, shapecolors );

Plot( LongZScore, "LongZScore", colorGreen, styleLine | styleOwnScale, -3, 3 );
Plot( ShortZScore, "ShortZScore", colorRed, styleLine | styleOwnScale, -3, 3 );
Plot( -2, "", colorBlue, styleLine | styleOwnScale, -3, 3 );
Plot( 0, "", colorBlue, styleDots | styleThick | styleOwnScale, -3, 3 );
Plot( 2, "", colorBlue, styleLine | styleOwnScale, -3, 3 );

////////// End //////////

```

Figure 2 – AmiBroker code for z-score system



Figure 3 shows the settings. No allowance is made for slippage. \$5 is deducted for commission for each trade. Trades are taken at the close of the bar that generates the action signal. (See Bandy, *Quantitative Trading Systems*, for a detailed discussion, including several techniques to accomplish this in practice.)

The screenshot shows the 'Backtester settings' dialog box with the 'General' tab selected. The settings are as follows:

- General settings:**
  - Initial equity: 100000
  - Positions: Long and short
  - Periodicity: Daily
  - Min. shares: 0.1
  - Min. pos. value: 0
  - ☐ Futures mode
  - ☒ Pad and align all data to reference symbol: SPY (turning this on may slightly change indicators if you have data holes)
  - ☒ Allow position size shrinking
  - ☐ Activate stops immediately (when turned on, stops are checked AFTER current bar signals)
  - ☒ Reverse entry signal forces exit
  - ☐ Allow same bar exit / entry
  - ☒ Use QuickAFL
- Defaults:**
  - Round lot size: 0 (zero means allow fractional # of shares)
  - Tick size: 0 (zero means no minimum change)
- Commissions & rates:**
  - ☐ commission table (Define...)
  - ☐ percent
  - ☒ \$ per trade: 5
  - ☐ \$ per share/contract
  - Annual interest rate: 0
  - Account margin: 100 (100 means no margin account)

Buttons at the bottom: Load, Save, OK, Cancel, Help.

Figure 3 – Trading system settings

## Walk Forward Testing

The system was tested using the walk forward technique. Figure 4 shows the walk forward settings. The period 1/1/2006 through 1/31/2012 was tested. In-sample length was set at one year. Out-of-sample length was set at one year. K-ratio, which rewards equity growth and rewards equity smoothness, was used as the objective function.

The screenshot shows the 'Backtester settings' dialog box with the 'Walk-Forward' tab selected. The settings are as follows:

- Mode:** Easy mode (EOD) (selected), Easy mode (Intraday), Advanced mode
- In-sample data:**
  - ☒ In-sample data
  - Start: 1/ 1/2006, End: 1/ 1/2007
  - ☐ Anchored
  - Last: 1/ 1/2012, ☐ Use today
  - Step: 1 year(s)
- Out-of-sample data:**
  - ☒ Out-of-sample data
  - Start: 1/ 1/2007, End: 1/ 1/2008
  - ☐ Anchored
  - Last: 1/ 1/2013, ☐ Use today
  - Step: 1 year(s)
- Optimization target:** K-Ratio
- Table:**

Mode	Start	End
IS	1/1/2006	1/1/2007
OOS	1/1/2007	1/1/2008
IS	1/1/2007	1/1/2008

Buttons at the bottom: Load, Save, OK, Cancel, Help.

Figure 4 – Walk forward settings

Figure 5 shows the walk forward summary results. As produced by AmiBroker, the report is quite wide. It has been edited and reformatted to ease interpretation. In this example, there are six walk forward steps, each an in-sample period (Mode == IS) followed by the associated out-of-sample period (Mode == OOS).

Mode	Begin	End	No.	Net Profit	Net % Profit	Exposure %	CAR	RAR	Max. Trade ...	Max. Sys % ...
IS	1/1/2006	1/1/2007	1676	1765.64	1.77	5.92	1.79	30.25	-4.51	-0.44
OOS	1/1/2007	1/1/2008	1	1803.06	1.80	7.07	1.82	25.71	-5.48	-0.57
IS	1/1/2007	1/1/2008	8155	2595.04	2.60	5.06	2.62	51.75	-2.94	-0.36
OOS	1/1/2008	1/1/2009	1	2994.64	2.99	4.46	3.00	67.27	-23.76	-4.48
IS	1/1/2008	1/1/2009	11077	4176.46	4.18	2.58	4.19	162.47	-2.13	-0.22
OOS	1/1/2009	1/1/2010	1	2839.81	2.84	7.95	2.86	35.92	-9.06	-1.67
IS	1/1/2009	1/1/2010	10270	2996.60	3.00	2.93	3.01	102.78	-7.97	-0.90
OOS	1/1/2010	1/1/2011	1	738.89	0.74	4.41	0.75	16.95	-6.82	-0.88
IS	1/1/2010	1/1/2011	1155	2274.62	2.27	3.21	2.30	71.76	-5.42	-0.72
OOS	1/1/2011	1/1/2012	1	1454.02	1.45	4.34	1.47	33.87	-14.14	-1.54
IS	1/1/2011	1/1/2012	2827	4106.35	4.11	8.36	4.15	49.70	-14.14	-1.39
OOS	1/1/2012	1/1/2013	1	0.00	0.00	0.00	0.00	N/A	0.00	0.00

Mode	Begin	End	No.	K-Ratio	# Trades	Avg Profit/Loss	Avg % Profit/...	Avg Bars Held	# of winners	% of Winners
IS	1/1/2006	1/1/2007	1676	0.3511	54	32.70	0.33	3.78	39	72.22
OOS	1/1/2007	1/1/2008	1	0.1072	58	31.09	0.31	4.09	37	63.79
IS	1/1/2007	1/1/2008	8155	0.3967	48	54.06	0.54	3.67	36	75.00
OOS	1/1/2008	1/1/2009	1	0.0506	30	99.82	1.00	4.83	22	73.33
IS	1/1/2008	1/1/2009	11077	0.5624	24	174.02	1.74	3.75	21	87.50
OOS	1/1/2009	1/1/2010	1	0.0905	34	83.52	0.84	7.09	25	73.53
IS	1/1/2009	1/1/2010	10270	0.3703	30	99.89	1.00	3.50	23	76.67
OOS	1/1/2010	1/1/2011	1	0.1069	35	21.11	0.21	4.20	21	60.00
IS	1/1/2010	1/1/2011	1155	0.2925	31	73.37	0.73	3.65	21	67.74
OOS	1/1/2011	1/1/2012	1	0.1138	34	42.77	0.43	4.26	20	58.82
IS	1/1/2011	1/1/2012	2827	0.2594	46	89.27	0.89	5.70	35	76.09
OOS	1/1/2012	1/1/2013	1	-1. #IND	0	N/A	N/A	N/A	0	N/A

Mode	Begin	End	No.	LongZLength	BuyLevel	BuySecond	ShortZLength	ShortLevel	ShortSecond
IS	1/1/2006	1/1/2007	1676	8	0	0.6	2	0	1
OOS	1/1/2007	1/1/2008	1	8	0	0.6	2	0	1
IS	1/1/2007	1/1/2008	8155	18	-0.7	1	3	0.3	0.7
OOS	1/1/2008	1/1/2009	1	18	-0.7	1	3	0.3	0.7
IS	1/1/2008	1/1/2009	11077	14	-2.2	0.2	12	0.2	0.7
OOS	1/1/2009	1/1/2010	1	14	-2.2	0.2	12	0.2	0.7
IS	1/1/2009	1/1/2010	10270	5	-0.2	1	10	1.4	0.3
OOS	1/1/2010	1/1/2011	1	5	-0.2	1	10	1.4	0.3
IS	1/1/2010	1/1/2011	1155	10	-0.1	0.9	2	2.5	0.5
OOS	1/1/2011	1/1/2012	1	10	-0.1	0.9	2	2.5	0.5
IS	1/1/2011	1/1/2012	2827	10	-0.2	0.7	20	0.1	0.4
OOS	1/1/2012	1/1/2013	1	10	-0.2	0.7	20	0.1	0.4

Figure 5 – Walk forward summary

## Key to the Columns of the Report

Mode:	IS for in-sample, OOS for out-of-sample.
Begin:	Beginning date for the period.
End:	Ending date for the period.
No.:	Run number from the optimization that scored highest.
	There is no significance to the number in the IS period.
	Only one OOS run was made in each step, so the OOS run number is always 1.

Net Profit:	Net profit for the one year period.
Net % Profit:	Net percent profit for the one year period.
Exposure %:	Average percentage of funds that were in a position at any given time.
CAR:	Compound Annual Rate of Return for the period.
RAR:	Risk Adjusted Return, which is CAR divided by Exposure %.
Max Trade % Drawdown:	Worst loss on an individual trade.
Max Sys % Drawdown:	Maximum drawdown relative to account balance.
K-Ratio:	The metric used to rank the alternatives.
# Trades:	The number of trades in the period.
Avg Profit/Loss:	The average profit per trade in dollars.
Avg % Profit/Loss:	The average profit per trade in percent.
Avg Bars Held:	Average number of days in a trade. Day of entry is counted.
# of Winners:	Number of trades that were winners.
% of Winners:	Percent of trades that were winners.
LongZLength:	Number of bars in the lookback period for long trades.
BuyLevel:	Z-score to signal a buy.
BuySecond:	Additional z-score needed to take additional long positions.
ShortZLength:	Number of bars in the lookback period for short trades.
ShortLevel:	Z-score to signal a short.
ShortSecond:	Additional z-score needed to take additional short positions.

## Interpretation of the Report

The fact that the best values of the parameters differ from one time period to the next is an illustration that the system requires periodic resynchronization.

The range of lookback used to compute z-score was 2 bars to 20 bars. In some periods the best range was 2 bars, in others it was 20, with most somewhere in the middle.

The range of z-score level required for a signal was 0.0 to 5.0. Z-score is the number of standard deviations by which the close differs from the mean of closes. If the data followed the Normal distribution, the z-score would reach levels above 2.0 less than 3% of the time, and below -2.0 less than 3% of the time. In most periods, signals were generated when the close was well within 1 standard deviation of the mean. The additional deviation required for a signal to take an additional position varied between 0.2 and 1.0 standard deviations.

Each of the periods – both in-sample and out-of-sample – is one year long. The number of trades per period ranged from 24 to 58—2 to 5 per month. Average number of bars held ranged from 3.5 to 7.1 as reported. Since the day of entry is counted as the first bar, a trade that entered on Monday and exited on Tuesday has a 2 bar hold.

The exposure ranged from about 3% to about 8%. All trades were taken with a fixed position size of \$10,000 from an account initially funded with \$100,000. (Trading with a fraction of the

account will be examined later.) If the system was always in exactly one position, the exposure would be reported as 10%.

The system was profitable in all six of the in-sample periods. This is not unusual, since each period had 250 daily bars and was being fit by logic that had six variable parameters.

The system was profitable in all six of the out-of-sample periods. This is an indication that the logic is robust and is recognizing patterns that preceded profitable trading opportunities; and that the synchronization between the logic and the data persisted beyond the in-sample period.

Note the performance of the system as the walk forward progresses. Each year is processed both as an in-sample period, where all combinations of parameters are searched seeking the best fit; and as an out-of-sample period, where only one test is made using the parameter values chosen as best. In some two-year sequences, the out-of-sample performance is better than the in-sample performance. That is due, in part, to differences in the opportunity for profit between the two periods. The important comparison is between the performance of a year when seen first as an out-of-sample year, followed by as an in-sample year. It will always be the case that in-sample performance is better. The ratio of out-of-sample to in-sample performance gives an indication of the efficiency of the system.

There have been six transitions from in-sample optimization to out-of-sample test. Each adds to the experience of keeping the system synchronized using the process that would be followed when the system was used for live trading. If this system was to be put into live trading, the parameter values used would be those of the final in-sample run.

### **Analysis of the Out-of-Sample Results**

The system was run manually six times, each time as a backtest using the date range and parameter values of one of the six out-of-sample periods. The first was 1/1/2007 through 1/1/2008 using values of 8, 0.0, 0.6, 2, 0.0, 1.0, obtained from row 2 of the walk forward report. This was followed by the remaining five which covered years 2008, 2009, 2010, 2011, and 2012. The individual trades were combined into a single best estimate set and used to determine system health and position sizing.

Figure 6 shows the percentage gained or lost by each trade in the sequence they occurred in the walk forward out-of-sample periods.

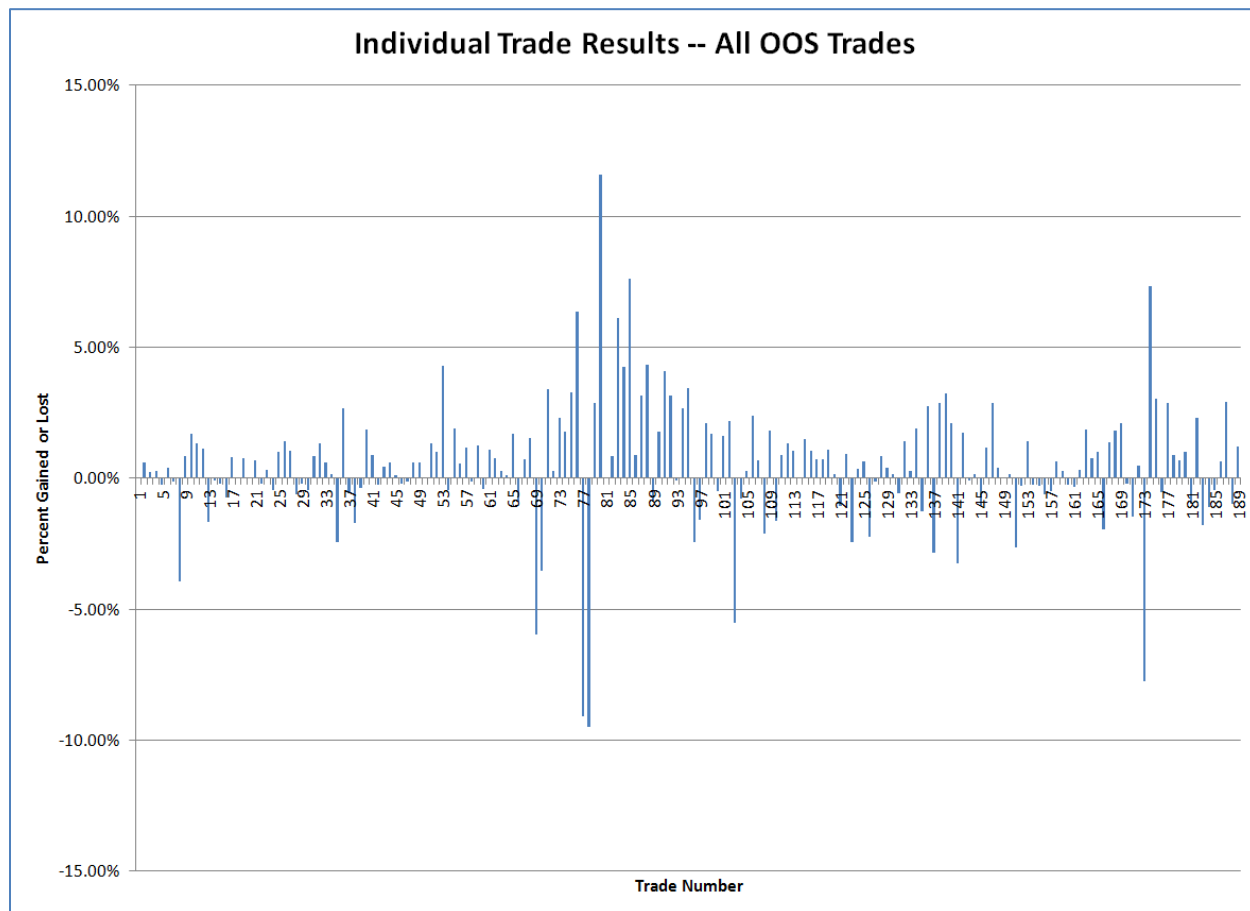


Figure 6 – Percentage gained or lost per trade

The equity curve that would have resulted if each trade was taken with a fixed size position of \$10,000 is shown in figure 7. Gain for the period 2007 through 2011 is \$9,782, all out-of-sample.

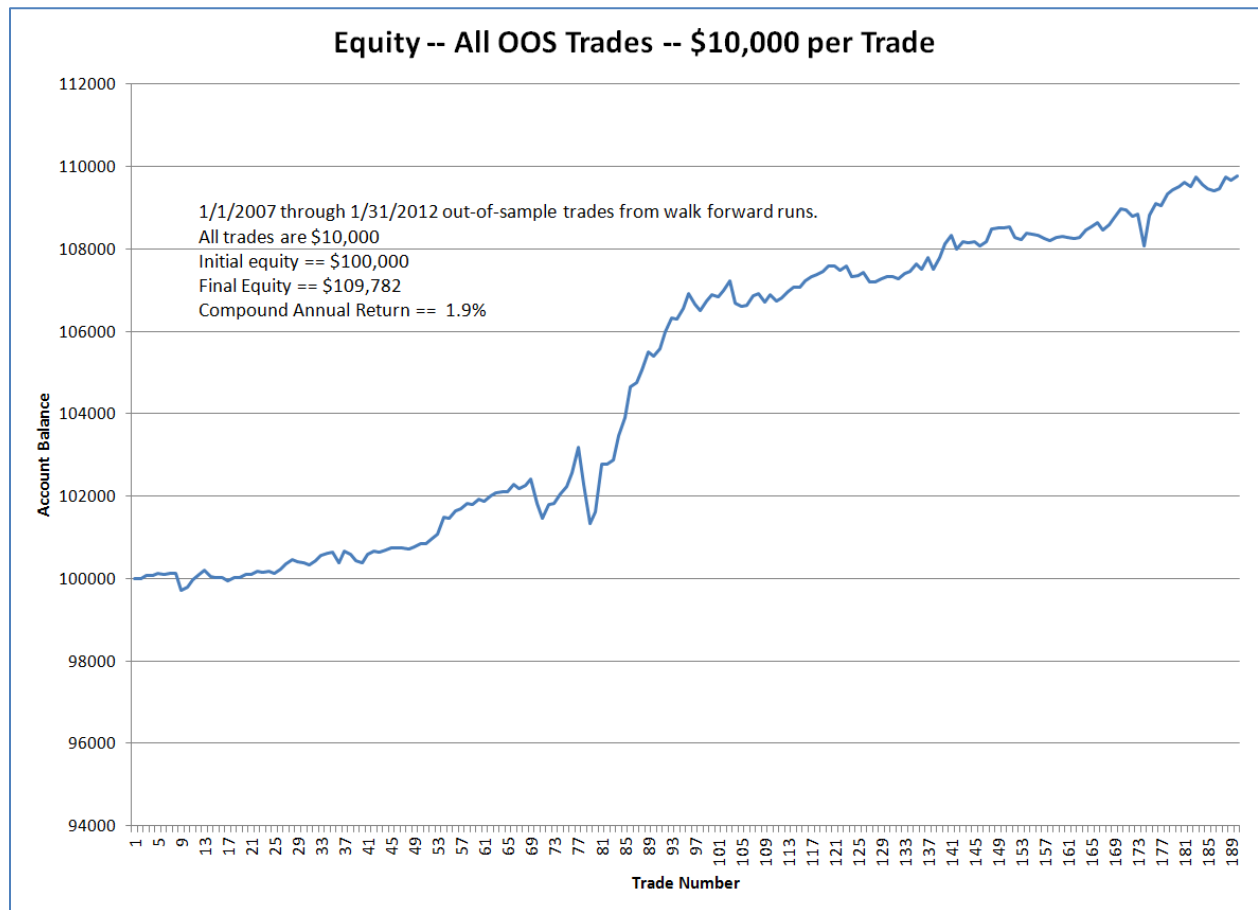


Figure 7 – Equity curve for \$10,000 positions

The equity curve that would have resulted if each trade was taken using all available funds (traded at a fraction of 1.00) is shown in Figure 8.

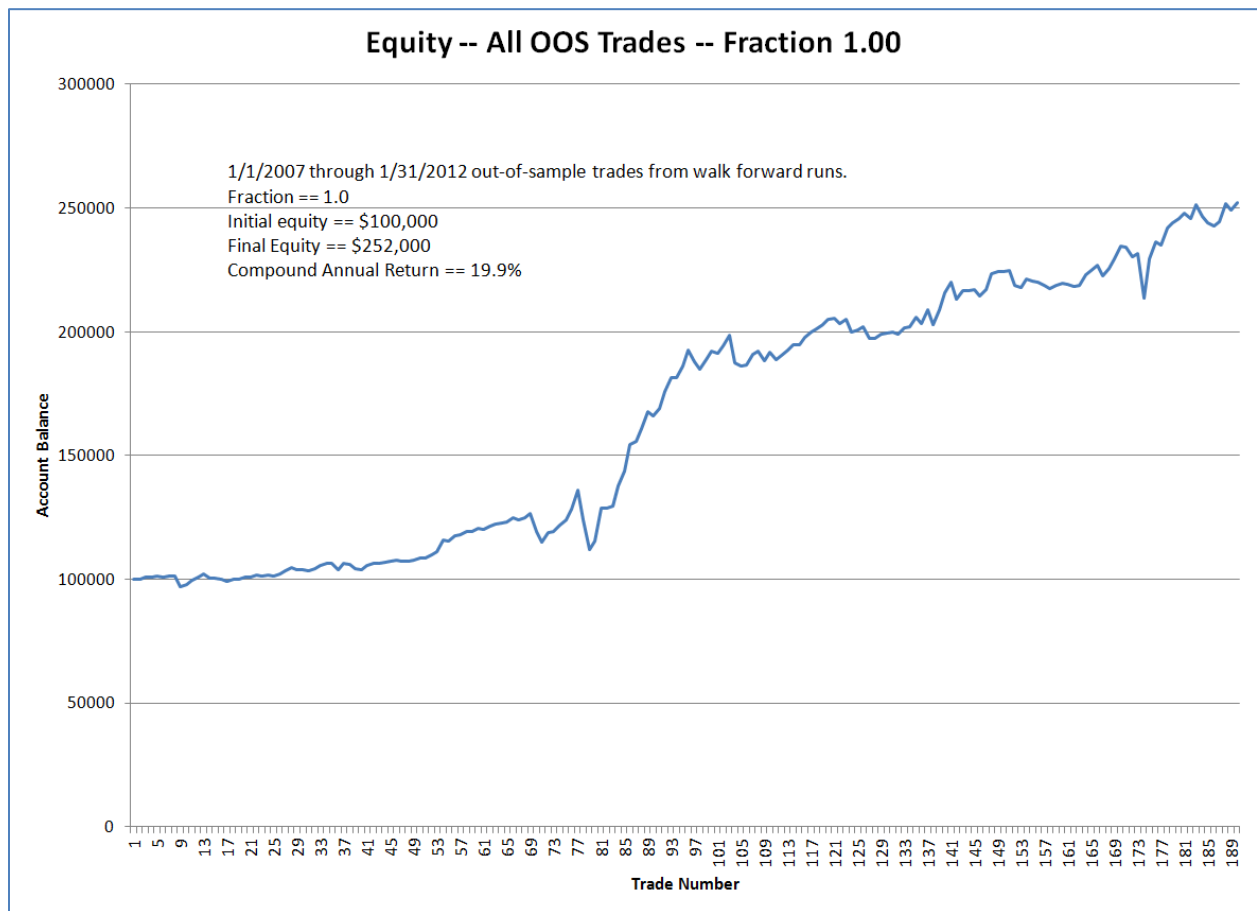


Figure 8 – Equity curve traded at fraction 1.00

Figure 9 shows the drawdown, expressed as a percentage of maximum equity to date, and the maximum drawdown. There was a significant losing sequence of two trades at about trade number 77, which caused a drawdown of about 18%.

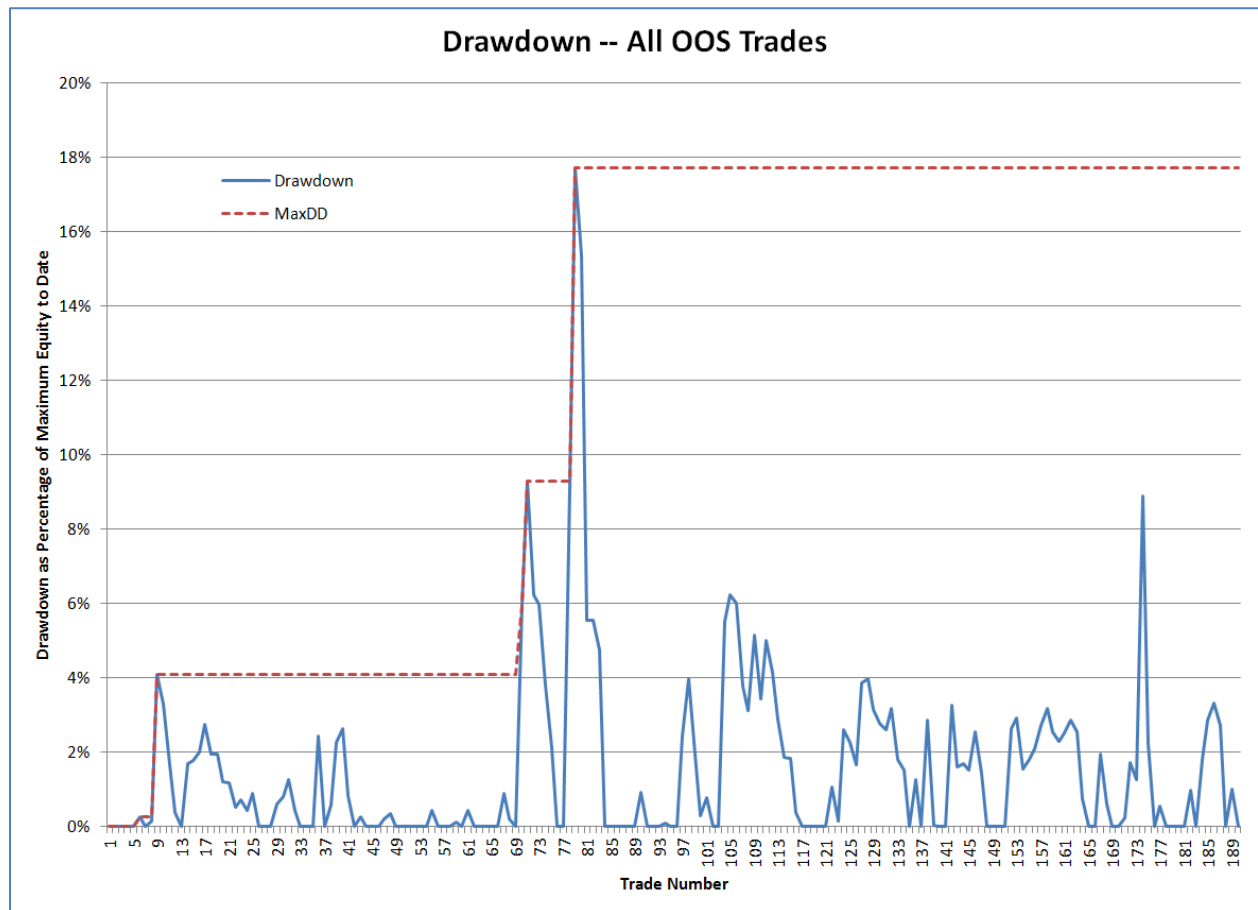


Figure 9 – Drawdown traded at fraction 1.00

## Further Analysis

Each of the walk forward steps consisted of one year in-sample followed by one year out-of-sample. It is interesting to examine the equity curve that would have resulted if any one of the one-year best parameter sets would have been used to generate signals for the entire period. The sequence of figures that follow show the date range used for the in-sample search between the red vertical bars and the equity curve that would have been achieved if the entire six year period used the values selected in that in-sample period. Among other things, this gives an opportunity to observe the performance of the system for the period immediately following the transition from development to trading.



Figure 10 shows the equity when 2006 was the in-sample year. Performance continued to be good for about one year into the out-of-sample period. The gain for the period 2007 through 2011 is about \$6,500. There was a serious loss in 2009 that has not yet fully recovered.

Figure 11 shows the equity when 2007 was the in-sample year. Performance was good and consistent for about six months, after which there was a serious drawdown followed by a very swift recovery. The gain for the period 2007 through 2011, one year of which is in-sample, is about \$6,000.

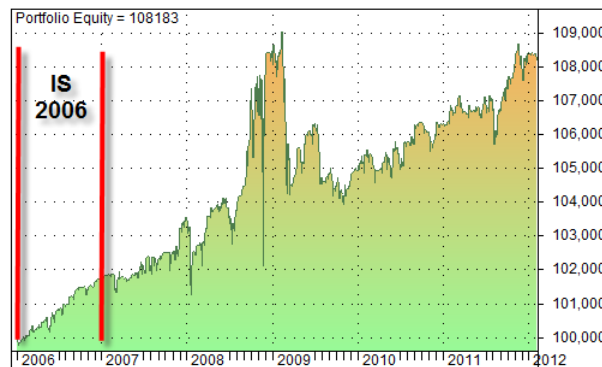


Figure 10 – 2006 is in-sample

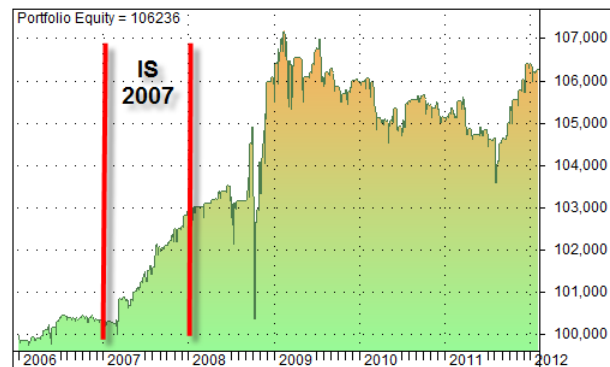


Figure 11 – 2007 is in-sample

Figure 12 shows the equity when 2008 was the in-sample year. The synchronization between the logic and the data was very close for the in-sample year. Out-of-sample performance was good, if inconsistent, for the next year, then flat. The gain for the period 2007 through 2011 is about \$7,000, most of which came from the in-sample year.

Figure 13 shows the equity when 2009 was the in-sample year. Out-of-sample performance continued to be good for about 18 months. The gain for the period 2007 through 2011 is about \$8,500.

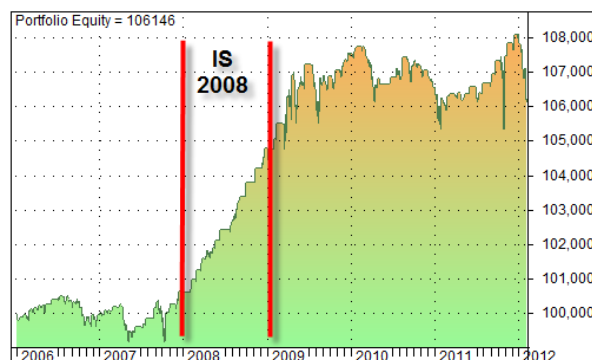


Figure 12 – 2008 is in-sample

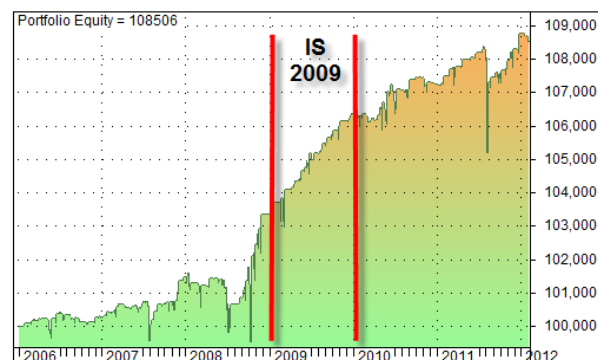


Figure 13 – 2009 is in-sample

Figure 14 shows the equity when 2010 was the in-sample year. Performance for the out-of-sample year 2011 is at about the same growth rate as the in-sample year, but has one sharp drawdown. The gain for the period 2007 through 2011 is about \$10,000, \$2,000 of which is in-sample.

Figure 15 shows the equity when 2011 was the in-sample year. The out-of-sample year will be 2012. There is a losing trade open at the end of 2011 that causes a drawdown in early 2012, but that trade has not been closed at the time this paper is being prepared, so there are no 2012 results to report. The gain for the period 2007 through 2011 is about \$9,000, \$4,000 of which is in-sample.

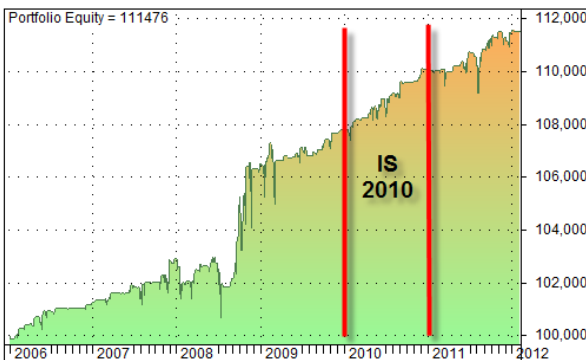


Figure 14 – 2010 is in-sample

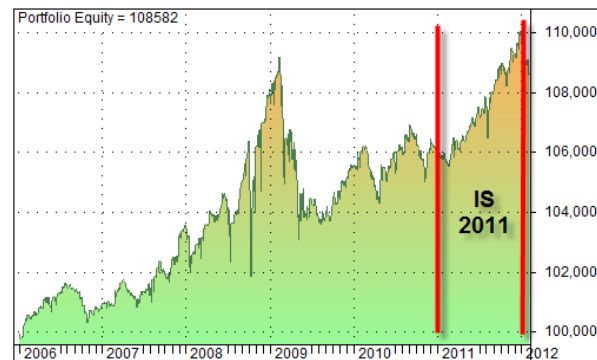


Figure 15 – 2011 is in-sample

## Observations

For all six steps, with the possible exception of the step when 2011 is in-sample, performance immediately following the in-sample period is satisfactory for at least six months and usually for a year or more. This tells us that the choice of a one-year in-sample period followed by a one-year out-of-sample period is reasonable, and that the system is quite robust.

No single set of parameter values chosen by the one-year optimizations had a six year out-of-sample performance superior to the concatenation of the six walk forward out-of-sample years. Referring back to Figure 7, a gain of \$9,782 was recorded for that period, all of which was out-of-sample.

## Determining System Health and Position Size

If a money manager likes this system, but is uncomfortable with the 18 percent drawdown that was experienced when traded at full fraction, she can use the Monte Carlo simulation technique to determine the position size that will produce the greatest equity growth while limiting drawdown to a level acceptable to herself and her clients.

She will begin by making some judgments:

- The time horizon. The number of months or years to project trades. There are about 37 trades per year, so a horizon of one to four years would be reasonable. The shorter the horizon, the fewer number of trades, and the more variation in the results. The longer the horizon, the greater the expected drawdown simply because expected drawdown increases with time. Say she decides on a two year horizon.
- The limit on drawdown she wishes to maintain. She might wish to avoid double-digit drawdowns, so that limit is set at 10% for this example.
- The confidence that the drawdown limit will not be exceeded. There is no certainty. But choosing position sizing so that there is 95 percent confidence is reasonably conservative.

The analysis is done using Monte Carlo simulation – a technique that enables developers to address problems too complex for equations. Monte Carlo simulation relies on repeated random sampling from a set of input values to estimate the mean and distribution of outputs. The best estimate we have of the gain or loss of future trades comes from the best estimate set of trades produced by the walk forward runs. As a first step, we need to estimate the distribution of maximum drawdown, at various position sizes, over that period.

The procedure is:

1. Decide the length of time being simulated. Two years, in this example.
2. Estimate the number of trades that will take place in that time. That will be 74 trades, based on the 61 month out-of-sample period having 189 trades.
3. Set the parameters for the simulation, such as the position size fraction.
4. Conduct many, say 1000, Monte Carlo runs. For each run:
  - a. Use sampling with replacement to select 74 trades from the best estimate set of trades.
  - b. Compute the account balance and drawdown after each trade.
  - c. Record performance metrics related to that run, such as final equity and maximum drawdown.
5. Create a distribution of the results of those 1000 runs.
6. Analyze the distribution.

Figure 16 shows the distribution of maximum drawdown based on 1000 runs, each a two year trading sequence, traded at fraction 1.00.

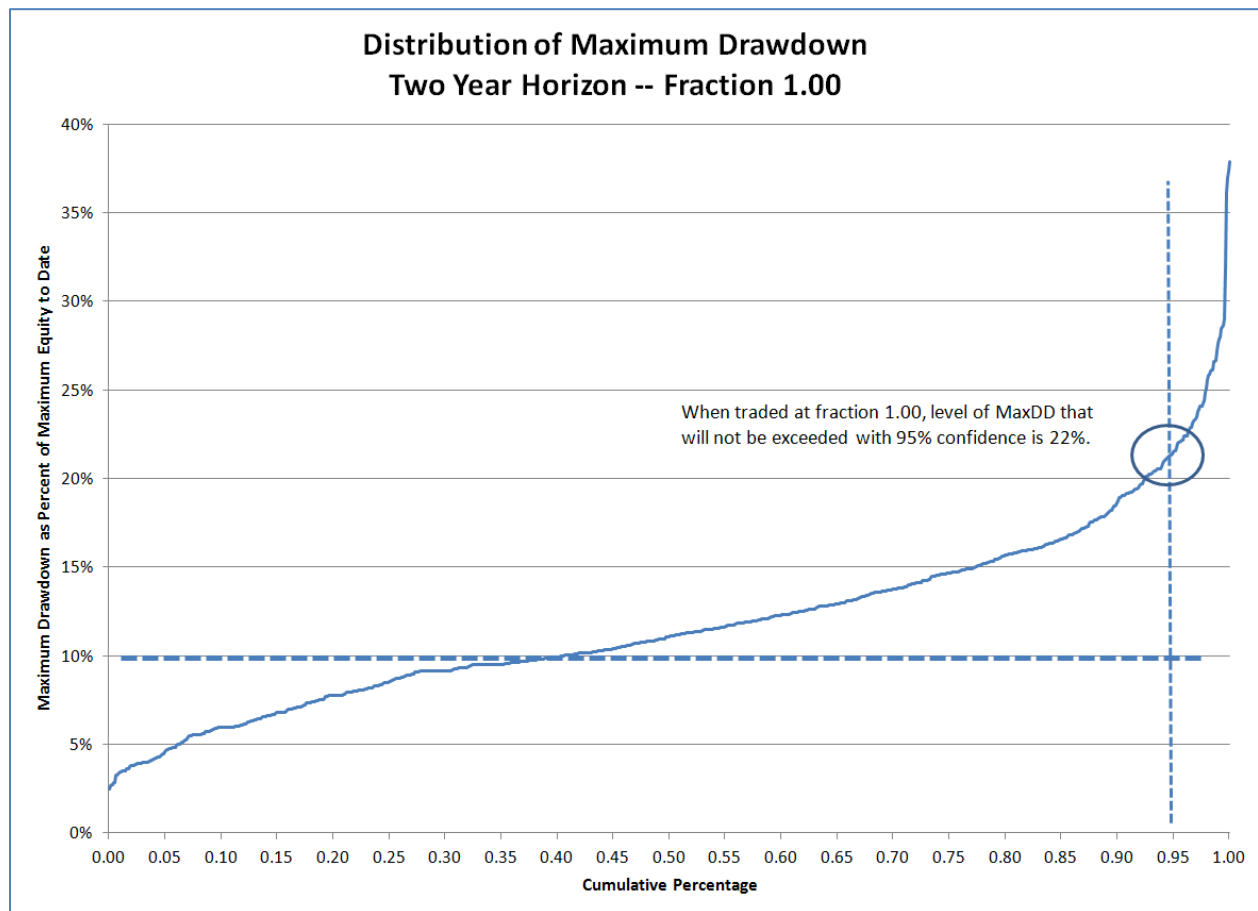


Figure 16 – Distribution of MaxDD traded at fraction 1.00

The manager wants to limit drawdown to 10% with 95% confidence. Both these levels are subjective choices of the manager. The limit to the maximum drawdown comes from the vertical axis of the chart; the degree of confidence comes from the horizontal axis.

As Figure 16 shows, there is about a 40% probability that MaxDD will be below 10% (the dashed horizontal blue line) and a 60% probability that it will be above 10%. The 95% confidence value (the dashed vertical blue line) shows there is a 95% probability that MaxDD will be below 22% and a 5% probability that it will be above 22%.

We need to know what the position size should be so that the distribution curve crosses the 10% horizontal line at the 95% vertical line. We learn that by running several more simulations, varying the position size from 0.10 to 1.00. Figure 17 shows the relationship between fraction and maximum drawdown. MaxDD 95 is the drawdown that will not be exceeded in 95% of cases. It is a conservative estimate of worst drawdown.

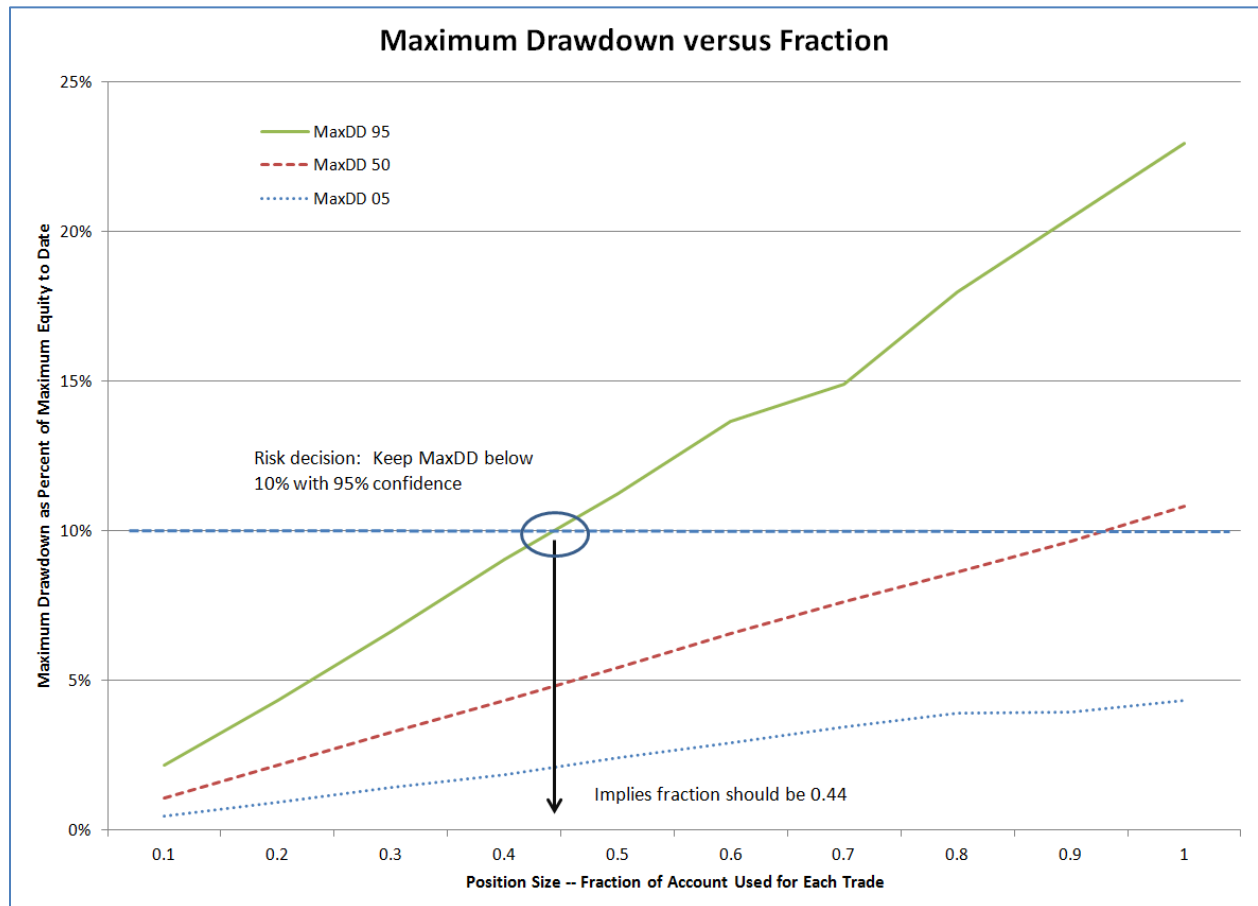


Figure 17 – Relationship between fraction and MaxDD

Interpolating, the MaxDD 95 line and the 10% line cross at about fraction of 0.44.

Based on the fraction chosen, the range of final equity can be estimated. See Figure 18. TWR 05 is the terminal wealth that will not be exceeded in 5% of cases. It is a conservative estimate of worst equity growth rate.

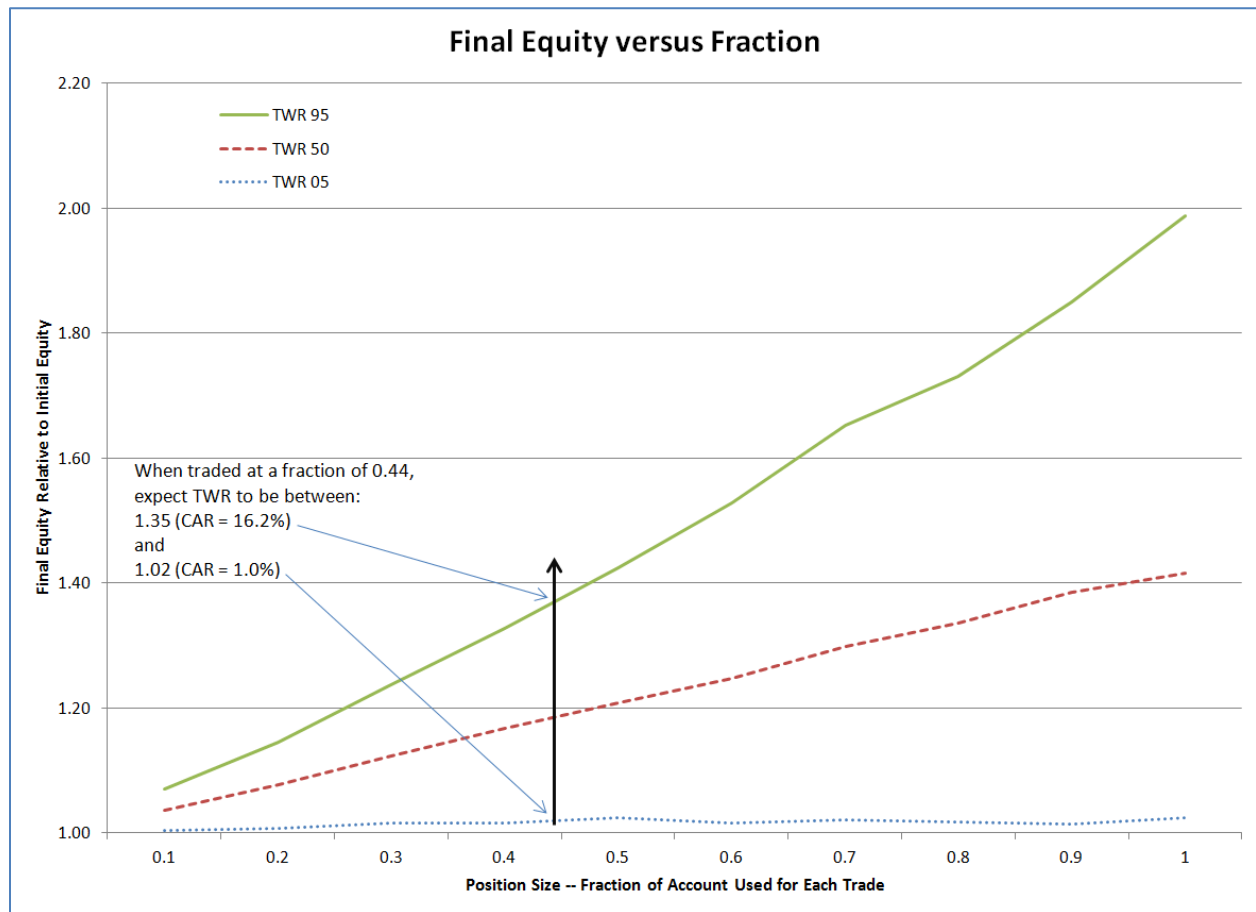


Figure 18 – Estimating final equity given fraction

### Expectations for Live Trading

The equity curve resulting from the out-of-sample trades of the walk forward runs (Figures 7 and 8) looked good – reasonably smooth and consistent. However, the single equity curve can be misleading. If the trades based on the distribution of the best estimate set occur in a different order, or if some of the trades are omitted, or some duplicated, the resulting equity curves may be better or worse than that single curve.

Given the limits set by the manager, maximum drawdowns of 10% or greater should occur with probability less than 5% when trades are taken using a fraction of 0.44 – that is, 44 percent of available funds. Refer back to Figure 18 and note the wide range of final equity between the 5th percentile and the 95th percentile – CAR values of 1.0% and 16.2%, respectively. The average is TWR of 1.19, giving CAR of 9.1%. Ten random trade sequences, each using a fraction

of 0.44, were constructed. Figure 19 shows the "straw broom" chart of their equity. The heavy black dashed line is the average, with a TWR after two years of about 1.16 and CAR of 7.7%.

The average of the ten is approximately the value expected, but the individual sequences vary widely. All are equally likely. Only one will happen, and we cannot predict which one.

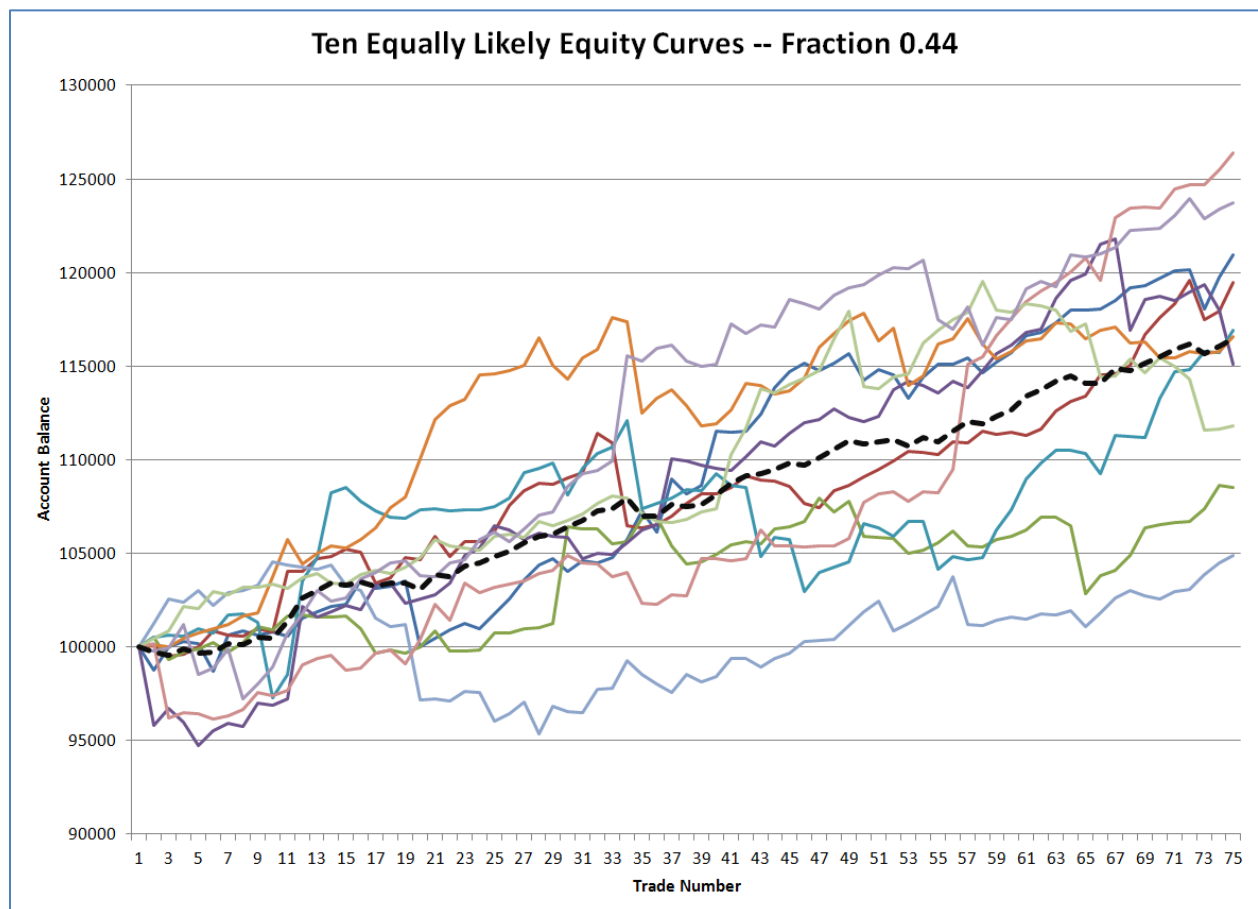


Figure 19 – Ten equally likely equity curves

Figure 20 shows the straw broom chart of the maximum drawdown. The heavy black line is the average.

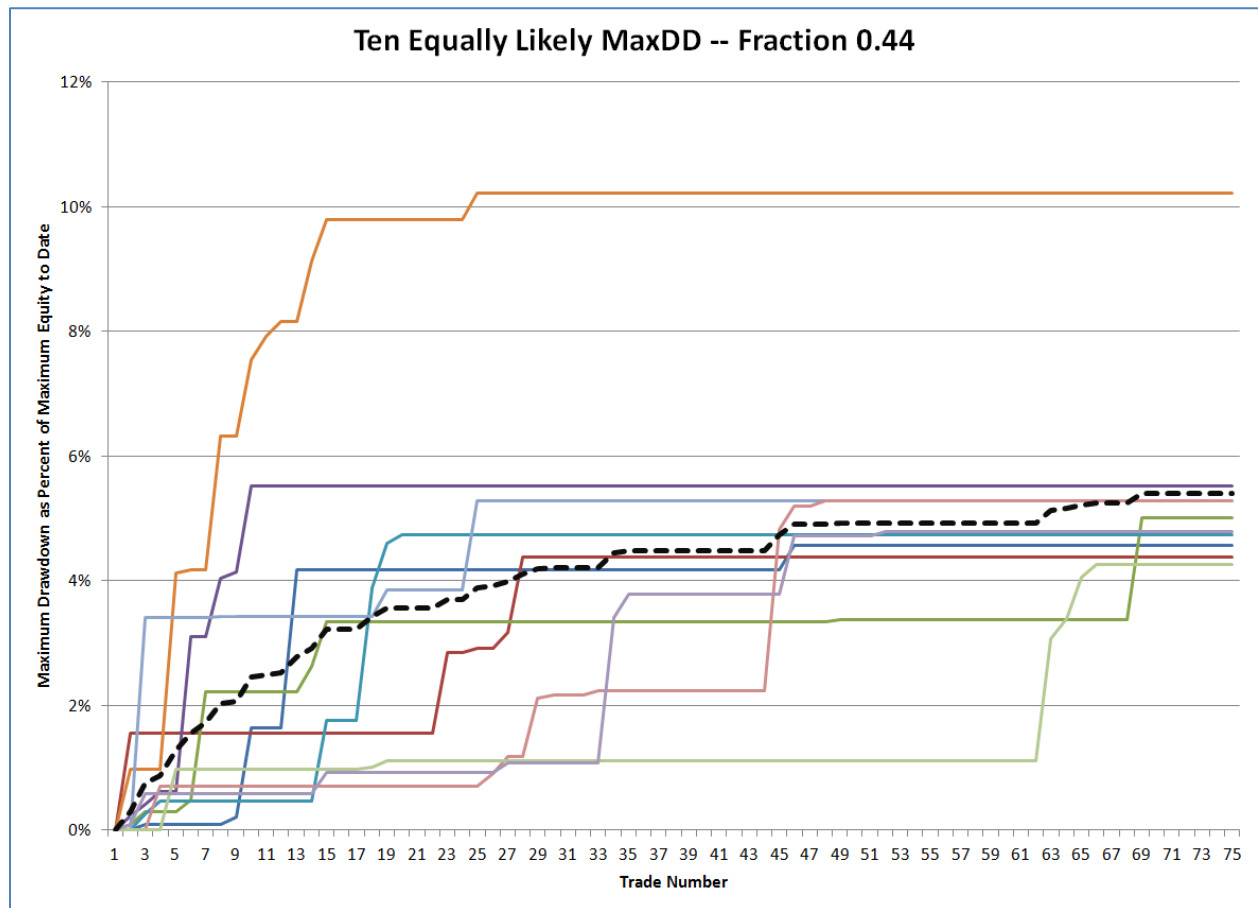


Figure 20 – Ten equally likely drawdowns

## Real-time Trading

As experience is gained from actual trades made by the system, or even paper trades, those trades should be added to the best estimate set and the position size recalculated periodically. By using a fixed length sliding window of the most recent trades to calculate position size and take advantage of the current degree of synchronization, system health can be continuously monitored. Deterioration of system health will be accompanied by lower recommended position size.



## Comparison with Buy and Hold

Figure 21 shows the equity curve resulting from taking a \$100,000 long position in SPY on 1/1/2007, the date the walk forward out-of-sample period begins. Final equity is \$88,774, a loss of 11.2%, giving a CAR of -2.3%. Figure 22 shows the drawdown.

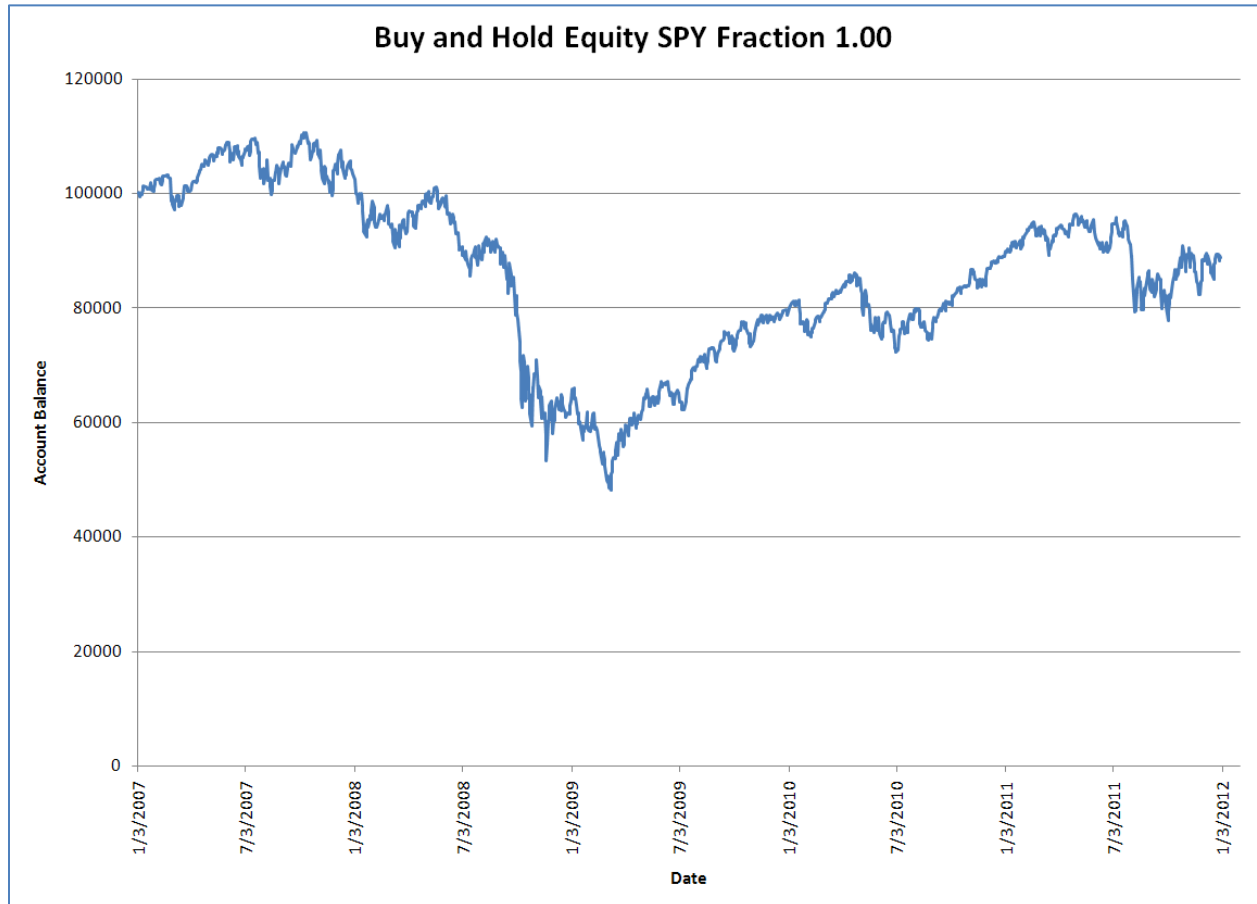


Figure 21 – Equity curve for buy and hold



Figure 22 – Drawdown and MaxDD for buy and hold

Figure 23 shows the equity curve from holding a position with 14% of available funds. 14% is the fraction that holds MaxDD below 10%, as shown in Figure 24.

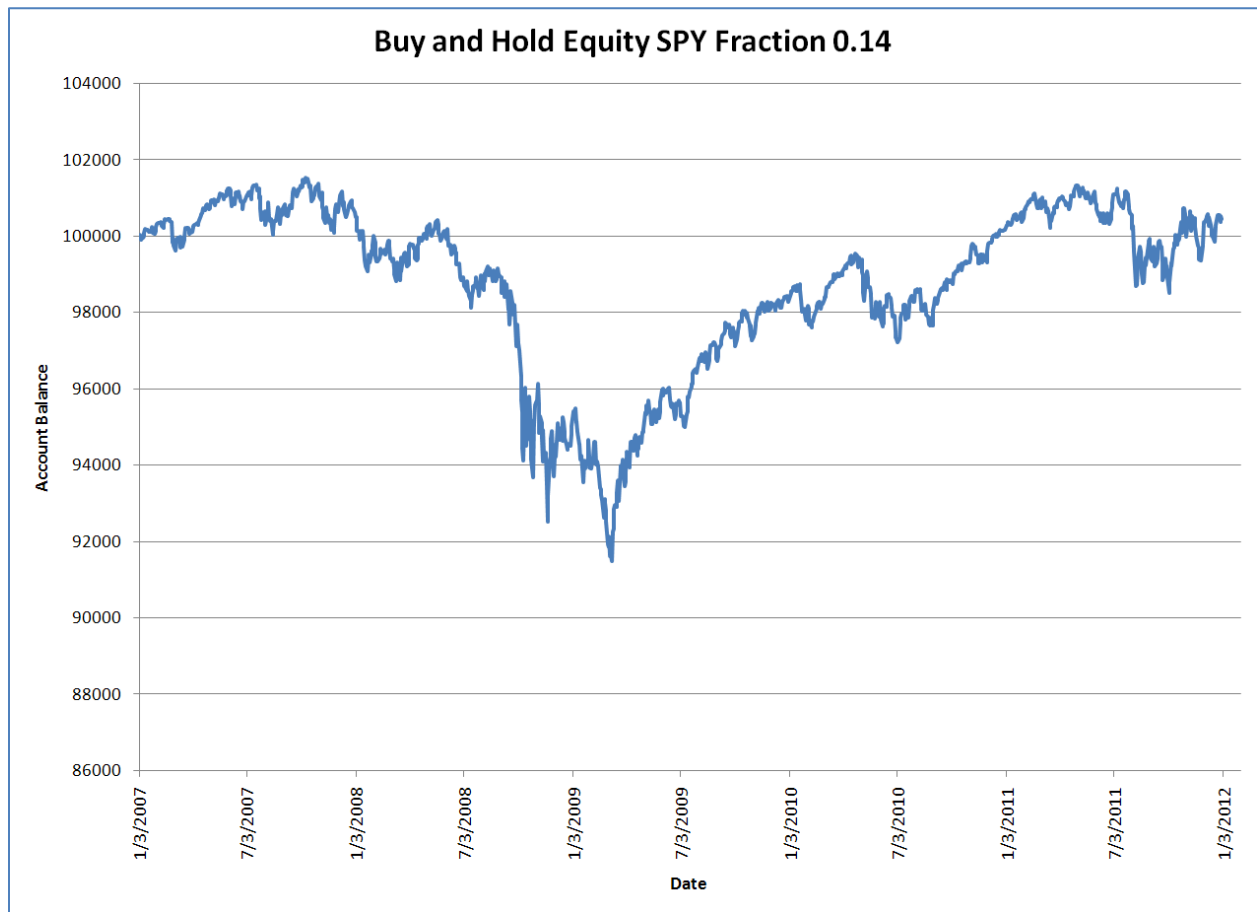


Figure 23 – Equity curve buy and hold at fraction 0.14

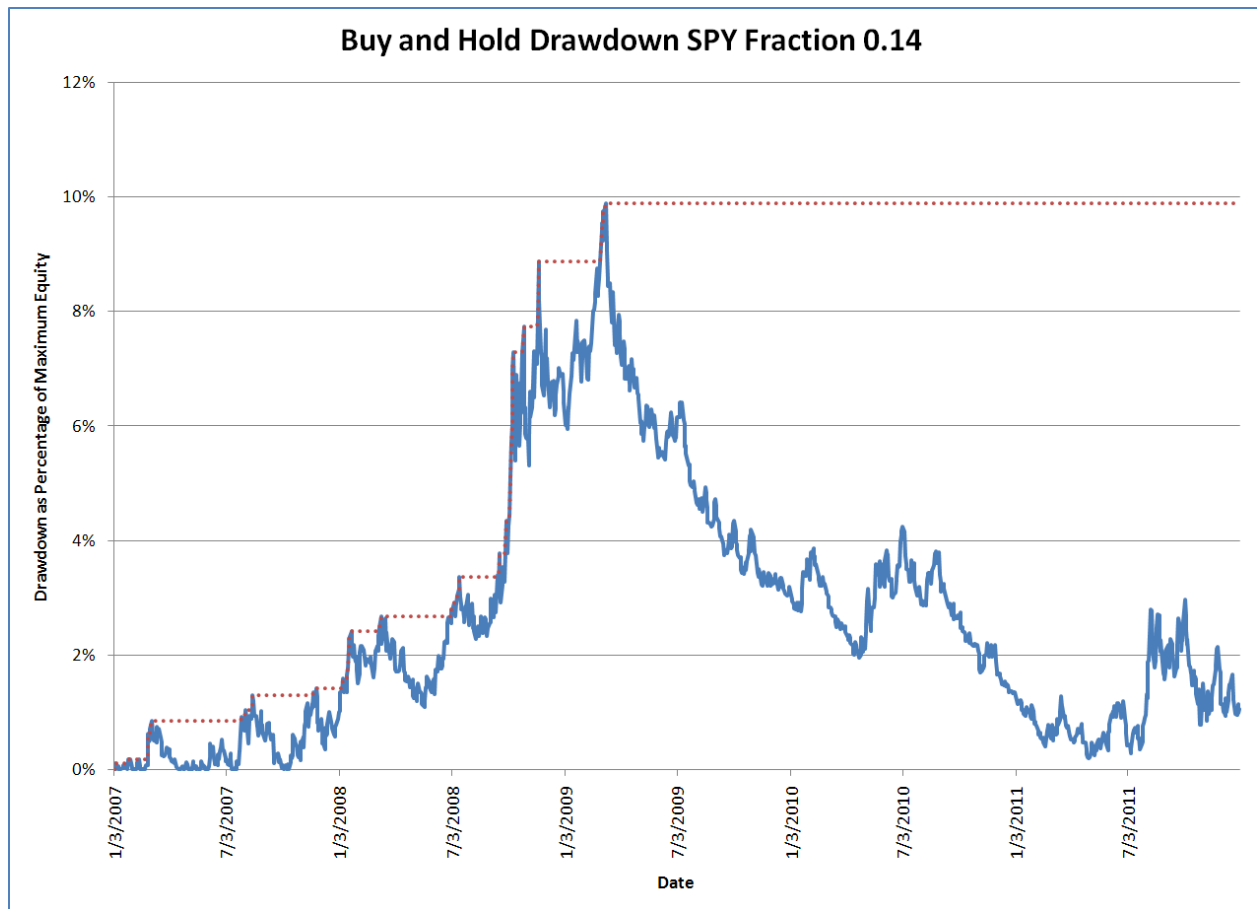


Figure 24 – Drawdown and MaxDD at fraction 0.14

Since there is no way to know the extent of the future decline in the price of SPY, use of a fraction such as 0.14 without that fraction being associated with a trading system, a best estimate set, and a distribution of maximum drawdown is not realistic. But it does illustrate a problem. A money manager who does not have a trading system and who has not defined a maximum loss level has a dilemma. Should he hold through steep drawdowns, hoping for recovery? Or go to cash at some point? If he does go to cash, at what point? And how does he determine that it is safe to reestablish the position?

Figure 25 shows the equity curve of a buy and hold position that exits to cash when the MaxDD reaches 10%. But it, alone, gives no method for determining what further action to take.

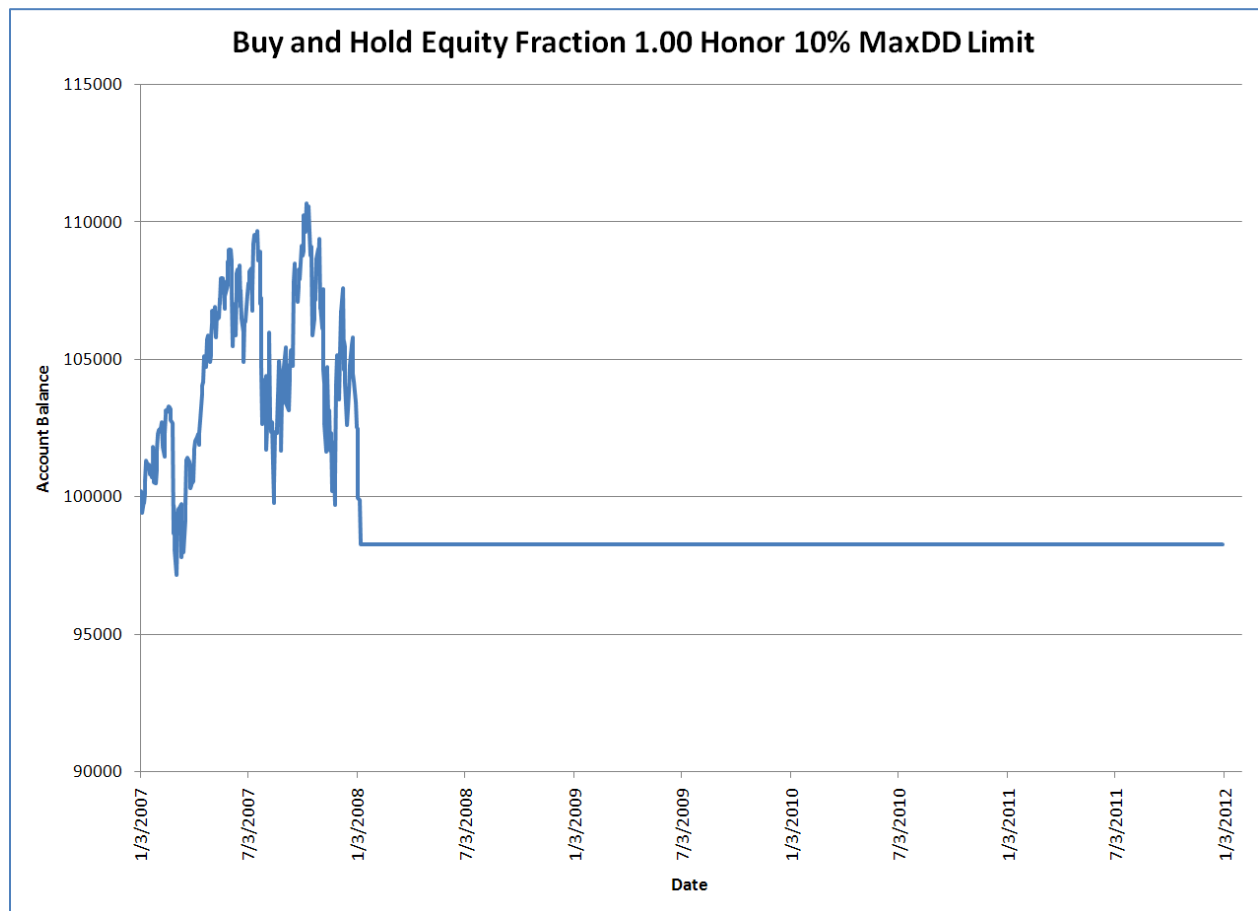


Figure 25 – Equity curve exiting at 10% drawdown

## Conclusion

It is important to have a well defined trading system. One that has passed tests of robustness and adaptability. One that you have a high degree of confidence in. Only by knowing the characteristics of the system can the health of the system be monitored and the reward and risk be established.

By choosing your own level of acceptable risk, you can determine the proper position size for use with your system to maximize account growth while limiting drawdown.

Remember: the correct position size for a system that is broken is zero.

##

## References

Bandy, Howard, *Quantitative Trading Systems*, Blue Owl Press, 2007, 2011.

– *Modeling Trading System Performance*, Blue Owl Press, 2011.

Vince, Ralph, *Portfolio Management Formulas*, Wiley, 1990

– *The Mathematics of Money Management*, Wiley, 1992

– *The New Money Management*, Wiley, 1995

– *The Handbook of Portfolio Mathematics*, Wiley, 2007

– *The Leverage Space Trading Model*, Wiley, 2009

## Website to order Dr. Bandy's books

[www.blueowlpress.com](http://www.blueowlpress.com)

## Blog about trading system development

<http://www.blueowlpress.com/WordPress/>

## Contact

[howard@blueowlpress.com](mailto:howard@blueowlpress.com)

##