Journal of Investment Strategies **2**(4), 59–81

The impact of stop losses on short-term countertrend trading strategies

Nicholas J. Libertini
361 Capital, LLC, 4600 South Syracuse Street, Suite 500,
Denver, CO 80237, USA; email: nick@361capital.com

(Received April 24, 2013; accepted July 5, 2013)

Stop losses are a common form of risk control for trading strategies, but there is a lack of empirical research regarding their effectiveness. Stop losses alter the statistical characteristics of a trading strategy, which makes it difficult to assess the tradeoff between improved risk control and return degradation. To evaluate a stop loss method equitably a systematic and quantitative approach is needed. This paper uses Ralph Vince's optimal fixed fractional position sizing combined with maximum drawdown to normalize the risk and return characteristics of a trading strategy. This risk and return normalization is then used to assess the effectiveness of stop losses as a risk control method for two short-term countertrend trading strategies. Specifically, a range of fixed percentage stops and a range of volatility scaled stops were evaluated. It was found that, due to the distinctive return path of short-term countertrend trades, stop losses were ineffective as a risk control system for these types of models.

1 INTRODUCTION

The primary objective of a stop loss is to automatically exit any losing trade that exceeds a predetermined amount in order to protect a trading system from large losses. Even though stop losses are utilized in all forms of trading, there is a lack of empirical evidence confirming their effectiveness.[1] Drawing substantive conclusions about the utility of stop losses is difficult because of the diversity of trading systems and stop loss methods. However, narrowing the research to a specific trading methodology should allow concrete conclusions to be drawn about the utility of stops. Focusing on

I would like to thank my 361 Capital colleagues, whose insights and research assistance have been invaluable: Jeremy Frank, Aditya Bhave, Brian Cunningham and J. B. Blue. I am also grateful to Kristine Lombardi for her tireless editing.

[1] Lei and Li also note this lack of research on stop loss effectiveness in their 2009 paper "The value of stop loss strategies".

momentum models, Kaminski and Lo (2008) show that momentum trading systems can benefit from stop loss policies.

This paper presents a quantitative approach for evaluating stop losses with an ultimate goal of measuring the effectiveness of stop losses as a method of risk control for short-term countertrend trading models. The mechanics of countertrend models appear well-suited to an additional layer of risk control. These models typically lack a mechanism for exiting escalating losses, and will instead wait for prices to revert to the mean before closing out a trade. It would seem that a stop loss designed to exit trades that fail to revert to the mean could improve the risk characteristics of these types of trading models.

This paper is organized as follows. Section 2 illustrates the ambiguity of comparing a trading strategy with and without stops. Section 3 outlines optimal fixed fractional leverage as a quantitative method for objectively evaluating the impact of stop losses. Results of stop losses applied to two short-term countertrend strategies traded on the S&P 500 index are presented in Section 4. Section 5 examines reasons for the failure of stop losses on short-term countertrend models and Section 6 concludes.

2 AMBIGUITY OF STOP LOSS EVALUATION

A stop loss can improve both a system's average trade loss and its maximum drawdown by closing out losing trades before they grow too large. When a stop is triggered, a guaranteed loss is realized in order to mitigate the risk that the loss could grow much larger. However, this increased control of downside risk likely reduces the upside return potential of a system because some losses are locked in that could have rebounded. This can make a system traded with stops appear less attractive than the original system from a return standpoint and difficult to evaluate using traditional performance metrics.

To illustrate the ambiguity of evaluating the impact of a stop loss, consider the returns of three hypothetical trading systems. System 1's trades were generated using a normal distribution with a mean trade return of 0.25% and a standard deviation of 2.00%. System 2 applied a 0.60% stop loss to System 1's trades with a sensitivity of 50% and specificity of 80%. System 3 applied a 0.55% stop loss to System 1's trades with a sensitivity of 50% and specificity of 65%.

The sensitivity and specificity proportions were used to emulate two drawbacks of stop losses. Sensitivity replicates the behavior that an end-of-day stop loss will not always exit every loss right at the loss threshold, resulting in a portion of losses that are still greater than the loss threshold. Specificity replicates the behavior that a stop loss will close out some trades at the stop loss threshold even though the trade's final return would have rebounded above the threshold. System 2's sensitivity implies that it had a 50% chance of successfully stopping out of any trade with a loss greater than

TABLE 1 Example systems stop loss results (in percent).

Statistic	System 1	System 2	System 3
GeoAvg trade return	0.26	0.23	0.17
StdDev of trade returns	2.02	1.66	1.61
Average trade loss	−1.55	−0.94	−0.85
Hit ratio	56.60	45.00	39.90
Max drawdown	−25.09	−17.52	−20.77
Optimally leveraged GeoAvg trade return	0.45	0.60	0.35

−0.60%. System 2's specificity implies that, for every trade that did not close below the −0.60% loss threshold, the stop loss still had a 20% chance (1 − specificity) of generating a false positive and incorrectly stopping the trade out for a −0.60% loss.

This framework assumes that nothing is known about the intratrade path of each trade and is therefore a simplified depiction of how stop losses affect a trading system. Although the sensitivity and specificity proportions are an imperfect method for modeling the interactions of a trade's path and a stop loss, the three systems still convey the ambiguity of interpreting stop loss results. Table 1 summarizes several traditional statistics for each system.

Systems 2 and 3 clearly lower the maximum drawdown and average trade loss of System 1. They also lower the geometric average trade return and the hit ratio. This makes the tradeoff between return reduction and risk reduction hard to weigh. Is a 0.64% reduction in average trade loss worth the decline in hit ratio from 56.60% to 45.00% for System 2? Is the improvement in maximum drawdown from −25.06% to −20.77% worth the 0.09% decrease in average trade return for System 3? Ultimately, a quantitative method is needed to compare this tradeoff of risk control and return to assess if a stop loss is truly effective. The quantitative method presented in this paper, optimal leverage, explicitly highlights that the value of the stop loss in System 2 is roughly 0.15% per trade, while the stop loss in System 3 detracts about 0.10% per trade. The procedure for calculating the optimally leveraged geometric trade return and using it to evaluate stop losses is outlined in the next section.

3 QUANTITATIVE EVALUATION OF STOP LOSSES USING OPTIMAL LEVERAGE

3.1 Optimal fixed factional position sizing

Optimal fixed fractional position sizing is a useful concept for understanding the tradeoff between risk and return in a trading system. Optimal fixed fractional position sizing considers both the return potential and the risk characteristics of a system to derive the optimal leverage that the system should be traded at. Rescaling the returns of a system by its optimal leverage normalizes its risk and return profile to a common scale so that it can be equitably compared with other trading systems. A popular form of fixed fractional position sizing is optimal Kelly (f_{Kelly}), which provides an analytic equation for calculating the optimal leverage that should be applied to a trading system in order to maximize its geometric growth per trade (Kelly 1956).

Optimal Kelly fixed fraction

$$f_{\text{Kelly}} = \left(\left(-\frac{\text{Avg.Win}}{\text{Avg.Loss}} + 1\right) \times P - 1\right)\left(-\frac{\text{Avg.Win}}{\text{Avg.Loss}}\right)^{-1}, \tag{3.1 a}$$

where P = probability of a winning trade.

Optimal Kelly leverage

$$\text{Leverage} = \frac{f_{\text{Kelly}}}{-\text{Avg.Loss}}. \tag{3.1 b}$$

Optimal geometric growth rate per trade

$$G(f_{\text{Kelly}}) = P \times \ln(1 + \text{Avg.Win} \times \text{Leverage}) + (1 - P) \times \ln(1 - f_{\text{Kelly}}). \tag{3.1 c}$$

To understand the mechanics of trading at optimal Kelly, assume a trading system has a 60% probability of winning on any given trade, with a return of +2.00% when it wins and a return of −1.00% when it loses. Optimal Kelly would dictate that the system should be leveraged 4000% because of its high positive expectation ((3.1 a) and (3.1 b)). Trading at optimal Kelly means that, on average, every trade the system takes will increase the trading capital by 14.83% (3.1 c), gaining 80% on a winning trade (4000% × 2.00%) and losing 40% on a losing trade (4000% × −1.00%).

Optimal fixed fractional position sizing is purely a mathematical relationship, expressing the optimal mix of leverage needed to exploit the full growth potential of a trading system. Although optimal fix fractional leverage is not something that most investors would use in practice, it is an informative measure of the impact of stop losses on a system. A successful stop loss will decrease the average loss size of a trading model, thereby increasing its optimal leverage (see (3.1 a)). In the example above, if a stop loss decreases the average trade loss to −0.50%, the trading system's optimal f_{Kelly} rises to 50%, which increases its optimal leverage to 10 000% and its growth rate to 38.19%. Furthermore, if the same 14.83% growth rate is desired as the original system, the leverage can be reduced to 1824% so that only the 9.12% of the trading capital is at risk when a trading loss occurs (1824% × −0.50%). In this case, optimal Kelly clearly highlights the benefit of the stop loss because the stop allows the system to be traded at the same 14.83% growth rate with much less risk (1824% leverage versus 4000% leverage).

Stop losses also affect the hit ratio of a system, turning some winning trades into losses by stopping them out before they can rebound. Using the former example, if the stop loss decreases the system's hit ratio from 60.00% to 44.10%, then the optimal f_{Kelly} drops to 30.13% and the geometric growth rate falls back to 14.83%. Even though the trading capital only depreciates by 30.13% on a losing trade, losing trades occur more frequently due to the lower hit ratio, which results in the same optimal growth rate as the original system. In this second case it can be concluded that the stop is not additive to the trading system because the optimal geometric growth rate of the system is the same with and without the stop loss.

Although useful for understanding the mechanics of fixed fractional position sizing, optimal Kelly is only applicable when a trading system's win size, loss size and probability of winning are constant across all trades or known exactly for every individual trade. In reality this is never the case. Ralph Vince's optimal fixed fraction (f_{Vince}) can be substituted for optimal Kelly to find the ideal geometric growth rate under more realistic conditions (see Vince 1990). Vince's optimal fixed fraction accounts for different sizes of wins and losses by using a system's historical return series to derive the position size that would have yielded maximum geometric return growth (Vince 1990). For this paper, f_{Vince} is combined with maximum drawdown to derive a trading system's optimal leverage (see (3.2c)). Maximum drawdown is considered by many to be an insightful measure of downside risk and it is a practical reference point for deciding how much leverage a system should be traded with.

Optimal Vince fixed fraction

$$f_{\text{Vince}} = \max_{0 \leqslant f \leqslant 1} G(f). \tag{3.2a}$$

Vince per-trade growth rate

$$G(f) = \left(\prod_{i=1}^{N} \left(1 + f \left(\frac{\text{trade return}_i}{-\text{largest loss}} \right) \right) \right)^{1/N}, \tag{3.2b}$$

where N = total number of trades

Optimal leverage

$$\text{Optimal leverage} = \frac{f_{\text{Vince}}}{-\text{Max drawdown}}. \tag{3.2c}$$

3.2 Outline of stop loss evaluation methodology

To make an accurate comparison between a system without stops and a system with stops, the optimal leverage of both systems should be calculated and used to scale each system's trade returns. The geometric average trade return can then be calculated for each scaled strategy and employed as an equitable metric for assessing whether stop losses added value to the system. This was the comparison done in the last row of Table 1 on page 61, which highlighted that the stops in System 2 were additive, while the stops in System 3 were detractive.

An additional layer of insight and robustness can be gleaned from this analysis by bootstrapping[2] each system's trade returns, calculating the optimal leverage for each bootstrapped sample, and then creating a distribution of the system's geometric average trade return. The bootstrapped distribution of the system with stops can then be compared with the bootstrapped distribution of the system without stops. For each stop loss tested in this paper, the trading system's original trades were paired with the stop loss augmented trades and the two return series were bootstrapped with replacement[3] to create 1000 unique samples. Each bootstrap sample consisted of a series of randomly selected trades from the original model and a series of the same randomly selected trades with the stop loss applied. In this way the series were paired, experiencing the same exact return path with the only difference being the change in the risk and return profile caused by the application of the stop loss.

Bootstrapping offers several advantages over a point comparison of one system's geometric average trade return with another. First, the calculation of optimal f_{Vince} relies on the largest loss of the system as a risk scale (see (3.2b)); the larger the maximum loss, the smaller the optimal leverage of the system. Bootstrapping with replacement harnesses this sensitivity by amplifying differences between systems with fewer large losses (ie, systems with stops) and systems that do not control outlier losses. This occurs because a system with fewer large losses will have a smaller probability of sampling one of these large losses for each bootstrap run, yielding a higher average optimal f_{Vince} across runs. Second, a successful stop loss should increase the similarity of a system's losses by capping losses larger than the stop threshold. If a system's losses are more homogeneous, this consistency will trickle down into the bootstrapped samples of its maximum drawdown making the denominator in (3.2c)

---

[2] Bootstrapping is a resampling technique that is used to estimate the properties of a statistic. Specifically, bootstrapping in this sense refers to the process of taking many random samples from the historical return series of a trading system and, for each sample, calculating a statistic of interest (ie, geometric average trade return). The variation in the values of the statistic across the random samples can then be used to infer properties of the statistic for the trading system.

[3] Bootstrapping with replacement creates a sample of the same size as the original sample by randomly drawing an observation from the original sample, adding it to the bootstrap sample and then replacing the sample and redrawing more samples until the sample sizes are equal. Replacing the observations after every random draw allows a single observation from the original sample to appear multiple times in the bootstrapped sample. If an entire sample is bootstrapped with replacement, roughly 33% of the original observations will be excluded from each sample, which creates variation between samples.

smaller for many of the bootstrap runs. Additionally, because the calculation of a system's maximum drawdown is path dependent, bootstrapping imparts a layer of robustness in the calculation by reordering the trade returns for each run.[4] Finally, because a stop loss caps losses and increases homogeneity it should reduce the variance in a system's geometric average trade return across bootstrap runs. A point estimate of a system's optimally leveraged average trade return offers no intuition about the variance of this average, while bootstrapping creates a distribution of the statistic which can be used to estimate uncertainty reduction.

Three statistics can be extracted from the paired bootstrap test for stop loss evaluation. First, the mean geometric average trade return across the bootstrap runs can be compared for the system with and without the stop loss. The statistical significance of the mean difference is derived using Welch's two-sample t-test[5] (Welch 1947) and bootstrapping each mean to estimate its standard error.[6] This comparison yields a fairly comprehensive measurement of the difference between the optimally leveraged geometric average trade return for the two systems, accounting for both the magnitude and variance of the statistic for both systems. A t-score of 1.65 or above would indicate at least a 95% probability that the stop loss improved the geometric average trade return of the trading system justifying a rejection of the null hypothesis that the stop loss is no better than the base model. Any t-score less than 1.65 fails to reject the null hypothesis.

The percentage of bootstrap runs in which the stop loss outperformed the base model can also be calculated, giving an estimate of the frequency that the stop loss will generate an improved risk and return profile. Finally, the difference between the stop loss's geometric average trade return and the base model's geometric average trade return can be calculated for every bootstrap run and used to create a ratio of the average positive difference to the average negative difference (difference ratio). The difference ratio is worth examining because a stop loss may underperform the base model by a small amount in many of the bootstrap runs, yet vastly outperform the base model in a few extreme cases. This would give the stop loss an option-like payoff and a very favorable difference ratio ($\gg 1$) might compel someone to implement the stop

---

[4] The historical returns of a trading system could be reordered and the total return and volatility of the system would remain the same, but the maximum drawdown would be different. This behavior makes drawdown calculations dependent on the observed path of the original return series.

[5] Welch's t-test is the version of the two-sample t-test used when the two sample variances are not equal.

[6] Fixed fractional optimization causes any bootstrap run that produces a return series with a negative total return to be suppressed to 0. Thus, the distribution of optimally leveraged geometric average trade returns is truncated on the left side of the distribution at 0, creating a nonnormal distribution. Bootstrapping the calculation of the mean overcomes this nonnormality and allows an estimate of the mean's standard error to be derived.

loss even if the statistical evidence is not definitive (t score < 1.65). This introduces a subjective aspect to the evaluation of stop losses because it requires one to balance the tradeoff of making a Type 1 versus a Type 2 error.

The paired bootstrap optimal leverage method of evaluating stop losses does have a few drawbacks. First, sampling trades with replacement implicitly assumes that there is no autocorrelation between trades. This assumption may not be accurate and could cause an understatement or overstatement of risk in any given bootstrap run because trades are sequenced naively rather than according to conditional probabilities derived from an autocorrelation structure. Some of this bias is offset by the pairwise comparison of each return series with and without stops, which ensures that the stop loss method is applied to the same return paths as the base model it is compared with. Another drawback is that the approach makes use of frequentist inference, which relies solely on the observed historical return series of a trading system to draw conclusions. A layer of Bayesian inference could be added to the analysis to overcome this limitation by weighting the bootstrap samples based on subjective beliefs about the future return paths that the trading system will encounter. Finally, optimal f_{Vince} and maximum drawdown may not represent the best measure of risk to reward for all trading systems. Other risk to reward measures, such as a system's Sharpe ratio, might provide better risk optimization in some cases.

4 STOP LOSSES ON SHORT-TERM COUNTERTREND TRADING STRATEGIES

4.1 Outline of trading strategies

Two short-term S&P 500 (SPX) trading strategies are explored in this paper: a ten-day Bollinger band (BB) strategy and a ten-day relative strength (RS) strategy. Both strategies were designed to invest contrary to short-term price trends in order to capitalize on the mean-reverting behavior of extreme price movements in equity markets. Specifically, the BB strategy measures the average price and the standard deviation of price of the SPX over ten-day periods. If the current price is greater than one standard deviation above the ten-day mean, the strategy will short the SPX until prices fall below the mean. If the current price is less than one standard deviation below the ten-day mean, the BB strategy will go long on the SPX until prices rise back to the mean. The ± 1 standard deviation threshold was chosen because this threshold is commonly used by traders of BB systems and it creates a trade profile that is representative of a typical countertrend model with moderate activity (average of twenty-seven trades per year).

The RS strategy uses the percentage of positive price changes relative to total price changes in the SPX over ten-day periods. If positive price changes have accounted for

70% or more of the total price movement over a ten-day window, the RS strategy will go short until the percentage of positive movement drops below 50%. If positive price changes have accounted for 30% or less of the total price movement, the strategy will go long until the percentage of positive movement rises above 50%. The 30%/70% threshold was chosen because of its common usage in industry and it is representative of a more selective countertrend model (average of thirteen trades per year).

Both strategies were traded on the SPX from December 31, 1989 to December 31, 2012 for a total of 5797 trading days. Each strategy was 100% long, 100% short or 100% cash, and short positions were rebalanced to 100% notional exposure at the end of each day.[7] Positions were implemented on the closing price of the SPX each trading day and no slippage or transaction costs were assumed. These strategies were explored in idealized states because only the performance difference between the base strategies and the strategies with stops were of interest. Table 2 on the next page summarizes the performance of the two strategies.

These two simplistic trading strategies do not exhaustively represent the performance of short-term countertrend trading. There are many other countertrend models with various statistical signatures that, when traded with stop losses, may yield different results than the results presented. Nonetheless, these simple models satisfactorily represent the theory of short-term countertrend trading and should offer valuable insight into the effect of stop losses on these types of models. Even though the strategies are mechanically very similar, the signal correlation between them was only 65.34% from December 31, 1989 to December 31, 2012, which injects a reasonable amount of diversity into analysis of stop losses on short-term countertrend models.

4.2 Outline of stop losses

Two types of end-of-day stop losses have been examined in this paper. The first stop explored was a fixed stop loss which exits trades if they exceed a fix loss level (ie, 2.00%). The second stop loss was a volatility scaled stop loss that uses the SPX's twenty-day average true range (ATR[8]) to adjust the stop loss level for recent market

---

[7] As a short position makes money its notional exposure drops, and as it loses money its notional exposure rises. For example, if the market rises 3% in one day, a -100% short position will increase to -103%, creating 3% leverage. A daily rebalance of short exposure will return the notional exposure to 100% at the end of the day. Daily rebalancing of short positions creates a trade profile that has characteristics similar to being short volatility. However, daily rebalancing of short positions does not affect short-term trading strategies as much as long-term strategies and it is a more conservative employment of leverage than not rebalancing.

[8] The ATR is calculated by finding each trading day's true range and then averaging the true ranges over the lookback window. A day's true range is defined as

$$\max(\text{High} - \text{Low}, \text{High} - \text{Last close}, \text{Last close} - \text{Low})/\text{Last close}.$$

TABLE 2 Trading Strategy Statistics from December 31, 1989 to December 31, 2012.

Statistic	BB	RS
Percent invested	74.68	59.10
Number of trades	624	306
Number of wins	436	200
Number of losses	188	106
Hit ratio (%)	69.87	65.36
Total return (%)	606.23	151.96
Annualized return (%)	8.87	4.10
Standard deviation (%)	16.19	14.33
Max drawdown (%)	−20.51	−24.81
Average gain (%)	1.50	2.00
Average loss (%)	−2.36	−2.76
Optimal f (%)	54.13	49.25
Optimal leverage (%)	263.89	198.48
Leveraged geometric average trade return (%)	0.69	0.49

volatility. Equity market volatility is highly transitory so it makes sense to have some method for adjusting a stop loss to accommodate changing market environments. Both stop losses were explored over a range of practical stop loss levels.

Only end-of-day stops are examined in this paper. Intraday stops present many testing and implementation issues that make them impractical as a risk control method for non-high-frequency trading systems. The major issue with intraday stops is the inability to consistently exit at the stop price because of speed and liquidity issues. Intraday stops may work as expected in low volatility environments with ample liquidity, but large intraday moves such as October 19, 1987, September 17, 2001, October 15, 2008 and May 6, 2010 cannot be protected against. In these instances, prices move too quickly through stop loss levels resulting in trade executions far away from the stop price.[9] Therefore intraday stops are ignored because they become ineffective when volatility rises and liquidity abates, which is the exact time when stops are truly needed for risk control.

[9] A *Wall Street Journal* article (Pilon *et al* 2010) following the May 6, 2010 US flash crash highlighted the risks posed by intraday stops. The article described how a trader's stop loss order was executed 16.30% below his stop price during the crash because of fast moving markets.

4.3 Results

Fixed percentage stops ranging from 1.00% to 15.00% and ATR stops ranging from 1.0 to 3.0 were investigated to evaluate the broad performance of stop losses on these two countertrend models. Table 3 on the next page and Table 4 on page 71 summarize the performance of ten different iterations of stop losses.

Examining the results in Table 3 on the next page and Table 4 on page 71, it appears that stop losses did not improve performance or risk of the short-term countertrend strategies. All of the tests returned a t-score of less than 1.65, providing insufficient evidence to reject the null hypothesis. Only the tighter stop losses ($\leqslant$ 3% for BB and $\leqslant$ 5% for RS) were consistently effective at reducing the average trade loss, but at the cost of a drastic reduction in hit ratio. Many of the stop losses degraded the hit ratio and increased the average loss size of the trading systems. The failure of stop losses to improve the return and risk profile of the models is also reflected in the difference ratios of each test, most of which were less than 1.00. This implies that even when a stop loss outperformed a base model on a bootstrap run it produced a smaller average positive difference in geometric average trade return than the average negative difference on other runs. Finally, the very low percentage of positive differences observed across bootstrap runs suggests that the stop losses failed to outperform the base model on a large variety of different return scenarios generated by resampling.

It may be argued that stop losses should only be implemented when market volatility is elevated because greater volatility presents amplified return risk for models that exploit price movements. To examine the impact of market volatility on the performance of stops for these two countertrend models, the rolling twenty-day standard deviation of the SPX was calculated from December 31, 1989 to December 31, 2012. Then the average of the twenty-day standard deviation was found over rolling one-year windows (252 trading days). All the trades from the countertrend models that occurred when the SPX's twenty-day standard deviation was above its 252 day average were classified as high-volatility trades and analyzed with the ten stops.[10] Of the 624 BB strategy's trades, 39.58% occurred when the SPX's twenty-day standard deviation was greater than its 252 day average. Of the 306 RS strategy's trades, 37.91% occurred at high volatility levels. Table 5 on page 72 and Table 6 on page 73 summarize the results of using stop losses with the high-volatility trades of the countertrend systems.

The results in Table 5 on page 72 and Table 6 on page 73 highlight that sixteen of the twenty stop losses did not statistically improve performance of the base model when trading at high volatility levels. Four of the stop losses did produce statistically

[10] The twenty-day rolling volatility on the day prior to a trade's first invested day was used as the volatility reference point.

TABLE 3 Bollinger band stop loss results.

Stop loss	Number of trades stopped out	Average trade loss (%)	Hit ratio (%)	GeoAvg trade return 95% CI (range, %)		*t*-score	Positive differences (%)	Difference ratio	Difference 95% CI (range, %)	
No Stop	0	−2.36	69.87	0.07	1.58					
Fixed percentage stops										
1.00%	272	−1.80	57.85	0.00	0.77	−27.22	3.40	0.22	−1.23	0.02
2.00%	185	−2.33	65.38	0.00	0.73	−27.33	1.60	0.27	−1.20	−0.01
3.00%	120	−2.37	68.11	0.03	1.13	−12.53	18.50	0.38	−0.87	0.20
4.00%	85	−2.52	69.23	0.01	1.00	−17.21	6.20	0.25	−0.88	0.06
5.00%	60	−2.44	69.23	0.04	1.23	−10.74	14.90	0.38	−0.70	0.14
10.00%	10	−2.56	69.87	0.01	1.23	−12.60	1.70	0.10	−0.58	−0.01
15.00%	4	−2.47	69.87	0.03	1.41	−6.05	10.40	0.27	−0.35	0.05
ATR scaled stops										
1.00	301	−1.59	54.01	0.00	0.94	−21.69	10.30	0.31	−1.14	0.15
2.00	203	−2.46	64.58	0.00	0.67	−33.67	0.10	0.15	−1.19	−0.05
3.00	137	−2.60	68.59	0.00	1.01	−19.25	5.20	0.25	−0.89	0.07

TABLE 4 Relative strength stop loss results.

Stop loss	Number of trades stopped out	Average trade loss (%)	Hit ratio (%)	GeoAvg trade return 95% CI (range, %)		t-score	Positive differences (%)	Difference ratio	Difference 95% CI (range, %)	
No stop	0	−2.76	65.36	0.00	1.84					
Fixed percentage stops										
1.00%	162	−1.63	48.04	0.00	1.17	−12.64	23.20	0.49	−1.36	0.45
2.00%	111	−2.29	57.52	0.00	1.11	−14.29	15.30	0.34	−1.32	0.20
3.00%	77	−2.62	62.75	0.00	1.38	−6.71	32.90	0.62	−1.05	0.51
4.00%	51	−2.67	64.38	0.00	1.76	−0.77	48.10	0.92	−0.86	0.66
5.00%	35	−2.75	64.71	0.00	1.57	−3.01	41.40	0.69	−0.80	0.47
10.00%	13	−2.96	65.03	0.00	1.32	−10.75	3.00	0.18	−0.80	0.01
15.00%	3	−2.87	65.36	0.00	1.74	−4.79	3.40	0.27	−0.38	0.01
ATR scaled stops										
1.00	174	−1.62	44.44	0.00	0.70	−22.36	5.00	0.38	−1.60	0.07
2.00	118	−2.22	56.86	0.00	1.61	−3.74	40.60	0.75	−0.97	0.63
3.00	85	−2.56	62.75	0.00	1.76	1.31	54.60	1.09	−0.64	0.62

TABLE 5 Bollinger band stop loss results at high market volatility.

Stop loss	Number of trades stopped out	Average trade loss (%)	Hit ratio (%)	GeoAvg trade return 95% CI (range, %)		t-score	Positive differences (%)	Difference ratio	Difference 95% CI (range, %)	
No Stop	0	−2.76	69.14	0.01	2.68					
Fixed percentage stops										
1.00%	121	−2.07	54.30	0.00	1.01	−27.88	12.90	0.57	−1.47	0.58
2.00%	89	−2.56	61.72	0.00	1.26	−22.85	17.10	0.67	−1.46	0.69
3.00%	64	−2.64	65.63	0.00	1.89	−5.12	38.30	1.04	−1.33	1.34
4.00%	48	−2.90	67.58	0.00	1.72	−9.18	34.90	0.85	−1.36	1.19
5.00%	36	−2.93	67.58	0.00	1.72	−8.71	33.80	0.92	−1.37	1.18
10.00%	7	−3.22	69.14	0.00	1.76	−11.64	27.70	0.94	−1.39	1.20
15.00%	3	−3.01	69.14	0.00	2.05	−2.89	40.00	1.16	−1.32	1.46
ATR scaled stops										
1.00	121	−1.98	53.52	0.00	1.55	−12.63	29.40	0.82	−1.39	0.97
2.00	86	−3.08	64.45	0.00	0.93	−28.58	13.50	0.51	−1.49	0.44
3.00	60	−3.23	67.19	0.00	1.28	−20.22	20.00	0.71	−1.47	0.79

TABLE 6 Relative strength stop loss results at high market volatility.

Stop loss	Number of trades stopped out	Average trade loss (%)	Hit ratio (%)	GeoAvg trade return 95% CI (range, %)		t-score	Positive differences (%)	Difference ratio	Difference 95% CI (range, %)	
No stop	0	−3.20	63.16	0.00	2.93					
Fixed percentage stops										
1.00%	78	−1.89	45.86	0.00	1.61	−11.61	25.90	0.96	−1.59	1.26
2.00%	56	−2.58	54.14	0.00	1.64	−12.09	26.10	0.87	−1.66	1.10
3.00%	40	−2.97	59.40	0.00	2.00	−5.14	36.30	1.08	−1.65	1.53
4.00%	25	−2.97	61.65	0.00	2.87	4.74	49.90	1.55	−1.54	2.42
5.00%	17	−2.93	61.65	0.00	3.09	5.81	52.50	1.57	−1.48	2.67
10.00%	9	−3.33	62.41	0.00	2.22	−4.30	33.70	1.31	−1.66	1.95
15.00%	2	−3.33	63.16	0.00	2.59	−0.26	37.70	1.61	−1.66	2.29
ATR scaled stops										
1.00	76	−2.06	45.11	0.00	1.29	−15.58	21.30	0.81	−1.71	0.90
2.00	52	−2.56	57.14	0.00	3.03	6.54	53.20	1.66	−1.51	2.63
3.00	36	−2.93	63.16	0.00	3.90	11.03	61.50	1.85	−1.36	3.54

TABLE 7 Bollinger band stop loss results across model variations.

Look back	Threshold	Fixed percentage stops (%)							ATR scaled stops		
		1	**2**	**3**	**4**	**5**	**10**	**15**	**1.00**	**2.00**	**3.00**
5	40/60	−38.02	−31.47	−17.59	−9.83	−8.43	−5.48	−1.73	−39.07	−27.59	−9.05
5	30/70	−36.11	−31.38	−15.12	−12.99	−8.76	−3.37	−2.71	−36.42	−22.39	−11.60
5	20/80	−23.40	−20.82	−7.89	−9.90	−7.59	−5.83	−4.69	−22.59	−23.59	−10.95
10	40/60	−33.37	−28.74	−7.53	−13.65	−12.12	−10.30	−6.91	−33.38	−22.40	−14.86
10	30/70	−28.30	−27.56	−12.30	−17.36	−10.48	−12.30	−5.95	−28.49	−31.84	−19.12
10	20/80	−24.13	−28.99	−24.57	−25.98	−15.92	−14.19	−3.43	−24.36	−35.39	−29.88
20	40/60	−21.69	−20.47	−2.52	**5.04**	−0.22	−7.54	−8.26	−21.97	−24.17	−3.75
20	30/70	−21.22	−17.27	**2.48**	0.39	−3.67	−9.85	−5.97	−21.35	−22.39	**9.97**
20	20/80	−17.83	−13.19	−5.51	−11.28	−6.85	−10.18	−6.64	−17.60	−15.77	−4.65

TABLE 8 Relative strength stop loss results across model variations.

		Fixed percentage stops (%)							ATR scaled stops		
Look back	Threshold	1	2	3	4	5	10	15	1.00	2.00	3.00
5	40/60	−30.21	−24.76	−16.08	−5.16	−5.89	−6.41	−3.35	−29.97	−23.34	−5.55
5	30/70	−23.56	−20.63	−9.65	−7.80	−8.18	−2.16	−3.50	−24.34	−17.61	−8.07
5	20/80	−26.00	−18.50	−12.26	−7.98	−8.11	−6.41	−3.79	−26.43	−22.18	−9.69
10	40/60	−28.79	−24.37	−16.52	−6.54	−6.86	−5.92	−2.10	−26.47	−20.66	**2.51**
10	30/70	−12.38	−14.55	−6.70	−0.81	−3.02	−10.65	−4.66	−12.60	−3.70	1.25
10	20/80	0.57	−0.94	−2.42	−3.73	−3.22	−0.80	−3.19	0.64	**1.98**	−4.15
20	40/60	−4.12	0.37	−9.82	−6.41	−0.77	−7.75	−7.93	−4.20	−4.11	−6.26
20	30/70	0.02	−3.78	−12.33	−7.18	−8.46	−6.89	−2.74	−0.03	0.25	**1.79**
20	20/80	**4.34**	0.97	1.01	−0.19	**3.04**	1.61	0.02	**4.33**	**7.89**	**9.35**

significant t-scores for the relative strength system. The percentage of positive differences was only slightly above 50% for these four tests, implying that the statistical significance was derived from the stop loss significantly outperforming the base model on a few bootstrap runs. For the other sixteen tests, most of the t-scores became even more negative for the high-volatility trades indicating the stop losses were even more degrading to the trading systems in these environments. While it seems intuitive that a stop loss would have a risk control advantage in high-volatility environments, these results imply that stop losses are detrimental to performance for short-term countertrend systems at increased market volatility levels.

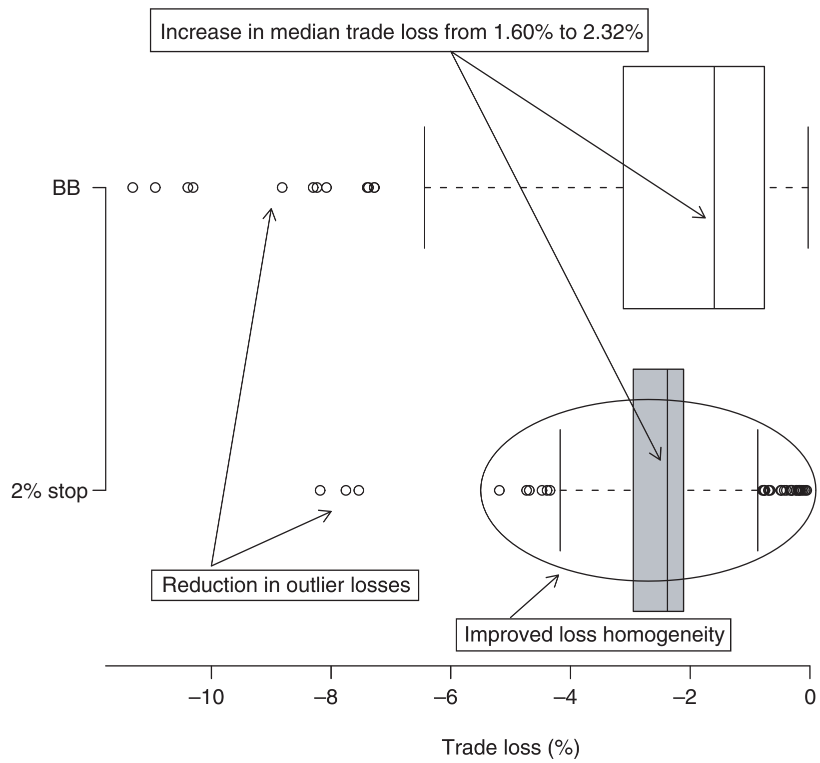
4.4 Sensitivity analysis

The findings thus far have shown overwhelming evidence that stop losses did not improve the risk profile of either of the short-term trading models. However, adjusting the input parameters of the two models may have an impact on the effectiveness of stops. To rule out the chance that all of the degraded performance was observed solely due to trading model inputs, both systems were evaluated with a range of input parameters using the same ten stop loss iterations. Specifically, the lookback window was varied from five days to twenty days for both models. For the BB system, 0.5 deviation, 1.0 deviation and 1.5 deviation thresholds were explored. For the relative strength system, 40%/60%, 30%/70% and 20%/80% trigger thresholds were explored. In all, nine model iterations were investigated for each model.

The sensitivity analysis results are summarized in Table 7 on page 74 and Table 8 on the preceding page. Each cell reports the t-score from the comparison of one stop iteration with one model input iteration. Of the 180 models tested with stop losses, 169 failed to reject the null hypothesis that stop losses are not additive to risk control. Eleven of the models had statistically significant t-scores (bold figures), but given the number of iterative hypotheses tested it is expected that some of these null hypothesis rejections were the result of spurious findings. Because ninety hypothesis tests were conducted for each of the trading models, a test should produce a t-score of at least 3.26 to reject the null hypothesis with 95% confidence.[11] This adjusted threshold produced only six rejected null hypotheses. Overall, the fact that 96.67% of the tests in the sensitivity analysis failed to reject the null hypothesis reinforces that stop losses are not an effective form of risk control for short-term countertrend models and that these findings are insensitive to the inputs of the models.

[11] When conducting multiple hypothesis tests the familywise error rate needs to be accounted for to adjust for the increased probability of making a false discovery. The Bonferroni procedure for familywise error rate correction stipulates that the confidence level (5%) should be divided by the number of hypothesis tests conducted (90). This correction results in a new confidence threshold of 99.94% and a corresponding t-statistic of 3.26.

FIGURE 1 Box plot of trade losses for a ten-day BB system with and without a 2.00% stop.



5 WHY STOP LOSSES FAIL ON SHORT-TERM COUNTERTREND STRATEGIES

The previous section provided compelling empirical evidence that stop losses are detrimental to short-term countertrend strategies. Figure 1 depicts the increase in risk when a 2.00% fixed stop loss is added to the ten-day BB model. This finding is counterintuitive because countertrend trading strategies let losses run, exposing trading capital to additional downside risk as they wait for prices to revert to the mean.

The gray box plot in Figure 1 illustrates that the 2.00% stop augmented the distribution of losses for BB model, but was not successful in reducing the risk characteristics of the trading system. The median loss size increased from −1.60% to −2.32% when using the stop, as indicated by the movement to the left of the black median line on the gray box plot. Furthermore, the number of losing trades increased from 188 to 216 when using the stop. These increases were a direct result of the stop loss exiting many of the system's trades prematurely. The stop loss did reduce the number of outlier losing trades of the BB system and it did increase the homogeneity of losses as indicated by the lower dispersion of outliers and the narrower width of the gray box plot. Nonetheless, these benefits were outweighed by the increase in the number of loses that were stopped out prematurely, which resulted in the hit ratio declining from 69.87% to 65.38%. A sizable decrease in a system's average loss is essential for risk improvement when employing a stop because the loss size reduction must offset the degradation of the system's hit ratio (see (3.1 a)). The fact that the 2.00% stop hardly decreased the BB system's average trading loss and degraded its hit ratio is strong evidence that it is not an effective form of risk control.

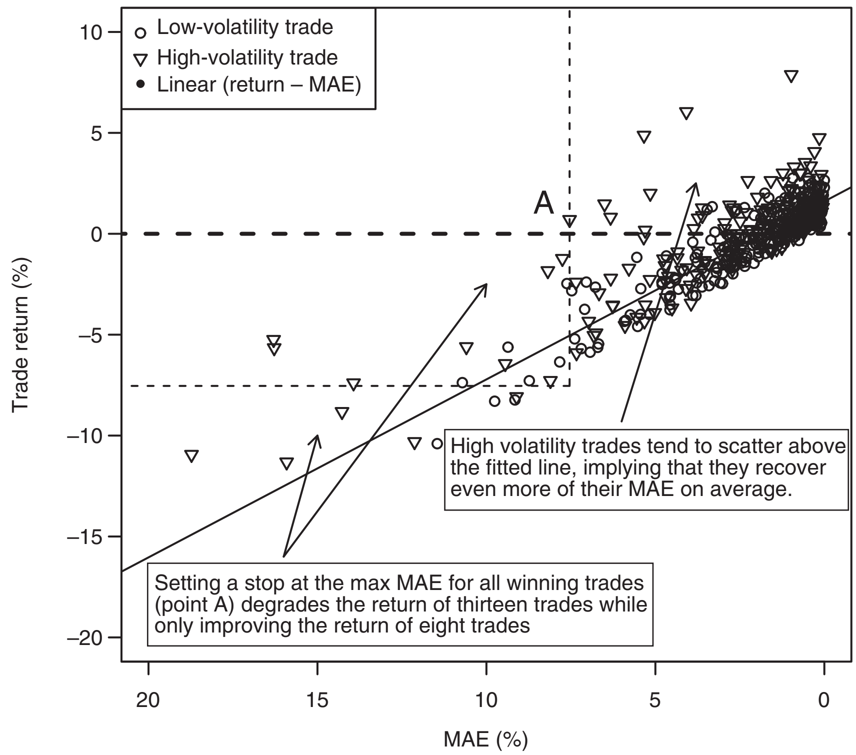
The return path of countertrend models is the primary reason why stop losses have a resounding negative impact on performance. As a loss moves against a countertrend model, the model's trade signal becomes stronger. For example, if a price falls one standard deviation below its mean, the BB system will go long, anticipating a price reversal. If price continues to decline to two deviations below its mean, the BB system will stay long, anticipating an even larger reversion to the mean. Therefore, stopping out of these trades de-risks the system when it is indicating that the probability of experiencing a reversion to the mean has increased.

To understand this phenomenon better, it is helpful to examine the relationship between a trade's maximum adverse excursion and its final return. Adverse excursion is price movement that generates a loss for an open trade. In the example above, the decline from one standard deviation to two standard deviations represents an adverse excursion incurred by the system. A trade's maximum adverse excursion (MAE) is the maximum loss that a trade experiences while it is open. For example, if a three-day trade has the following total return path: +1.00% total gain, −3.00% total loss, −0.50% total loss, the trade experienced an MAE of −3.00% before bouncing back to a −0.50% loss. The principal reason that stop losses are damaging to countertrend models is that a trade's final return is, on average, significantly better than its MAE for these types of trading system.

The scatter plot in Figure 2 on the facing page illustrates this point. For the trades taken by the BB system, the final trade return (y-axis) was plotted against the trade's maximum adverse excursion (x axis) and a robust linear regression[12] was fitted

[12] A robust linear model using the close-by maximum likelihood estimation method was constructed for the 434 BB trades that had nonzero MAE from December 31, 1989 to December 31, 2012. A robust linear model was used because they are less sensitive to outliers in both the explanatory and response variables. Several outliers were observed in the trade data due to extreme single day moves in the SPX over the simulation history.

FIGURE 2 Maximum adverse excursion versus trade return for Bollinger band system.



Trade return = 1.60% + 0.88 × MAE. R^2 = 69.91%.

between the two variables. The regression model (trade return = 1.60% + 0.88 × MAE) yielded an R^2 of 69.91% emphasizing the strong relationship between a trade's final return and its MAE. The regression equation highlights that a stop loss would have to be placed at least 1.60% above a trade's MAE to have a potential positive impact on risk reduction for the BB model. Applying the regression equation, a trade from the BB system that experiences an MAE of −2.00% will, on average, bounce back to a −0.16% loss before it is closed out (1.60% + 0.88 × (−2.00%)). In this example, a stop loss would have to be placed at −0.15% to improve the performance of trades that incur a −2.00% MAE. If the stop is placed below this level, it will lock in a larger loss than the trade would have naturally experienced, increasing the trading system's risk and degrading performance.

However, a −0.15% stop loss is impractical because it will stop out almost every trade that is not immediately profitable. This presents a major problem for the BB trading system. Seventy percent of the system's trades had a nonzero MAE, which means that such a tight stop would replace roughly 57%[13] of the system's positive trades with a small loss. It is characteristic for a short-term countertrend model to incur some degree of adverse excursion in the majority of its trades, and that is why tight and medium range stops degrade the performance of these types of models.

Given that tight stop losses cut off a large portion of winning trades for countertrend systems, a stop loss could instead be placed just below the max MAE for a system's winning trades in an effort to circumvent this impact. On the surface this seems to be a clever way to control risk while not limiting upside because none of the winning trades are stopped out. However, this method of stop loss placement is still detrimental for countertrend systems. Losing trades generally have larger MAEs than winning trades, but still recover a sizable portion of their MAE before they are closed out. A stop loss placed just below the max MAE of a system's winning trades will increase the loss size of many of the system's losing trades, degrading performance. Ultimately, the perpetual loss inflation at almost every stop loss level is why stops are ineffective as a form of risk control for short-term countertrend systems.

6 CONCLUSION

Stop losses are designed to improve the risk control of trading systems by limiting large losses and containing drawdowns. The hit ratio of a system declines when employing stops because some trades are closed out before they have a chance to recover and generate a profit, but an effective stop loss should more than make up for the winning trades it misses by greatly improving downside risk.

Vince's optimal fixed fractional leverage is an elegant way to accurately compare a system with and without stops (Vince 1990). If a stop loss has truly improved the risk characteristics of a trading system, then the system's bootstrapped largest loss and bootstrapped maximum drawdown will be less than the original system. These improved risk characteristics will be accounted for by an increase in the system's optimal leverage, allowing it to be traded more profitably than the original system. If a trading system can be traded at higher leverage with stops it is clear evidence that stops have improved its mathematical expectancy.

This research examined the effectiveness of utilizing stops with two countertrend trading models. When applying both fixed stops and volatility scaled stops, it was apparent that almost all of the stop losses degraded performance. The optimal leveraged geometric average trade return for most of the systems was statistically worse

[13] The BB system had a hit ratio of 69.87% and 70% of its trades had nonzero MAEs. Thus, 30% of its trades were winners that experienced no adverse excursion resulting in zero MAE. That leaves 30% of the trades that were losers and 40% of the trades that were winners; 40%/(30% + 40%) = 57.14% of winners had nonzero MAEs.

when using the stop loss for risk control. This occurred because as a loss moves against a countertrend model, the model's trade signal actually gets stronger. Stopping out of trades that continue to move against these models de-risks them when the probability of experiencing a reversion to the mean is highest. Ultimately, this causes a stop to lock in many losses that would have bounced back to smaller losses had they been allowed to run their course. The negative impact of de-risking was illustrated in the results of the systems traded with stops. The risk of each system increased when it was traded with stops, becoming trapped in larger drawdowns and cutting short many potentially profitable trades. This finding, although counterintuitive, implies that stop losses should not be used with short-term countertrend models.

REFERENCES

Kaminski, K. M., and Lo, A. (2008). When do stop-loss rules stop losses? Working Paper, Swedish Institute for Financial Research (May).

Kelly Jr., J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal* **35**(4), 917–926.

Lei, A. Y. C., and Li, H. (2009). The value of stop loss strategies. *Financial Service Review* **18**, 23–51.

Pilon, M., Blumenthal, K., and Zweig, J. (2010). When "stop loss" trades backfire on investors. *The Wall Street Journal*, May 15.

Vince, R. (1990). *Portfolio Management Formulas*. John Wiley & Sons, New York.

Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika* **34**, 28–35.