CHAPTER 13

Choosing Stop-Loss Points

It was the same with all. They would not take a small loss at first but had held on, in the hope of a recovery that would "let them out even." And prices had sunk and sunk until the loss was so great that it seemed only proper to hold on, if need be a year, for sooner or later prices must come back. But the break "shook them out," and prices just went so much lower because so many people had to sell, whether they would or not.

—Edwin Lefèvre

The success of chart-oriented trading is critically dependent on the effective control of losses. A precise stop-loss liquidation point should be determined *before* initiating a trade. The most disciplined approach would be to enter a good-till-canceled (GTC) stop order at the same time the trade is implemented. However, if the trader knows he can trust himself, he could predetermine the stop point and then enter a day order at any time this price is within the permissible daily limit.

How should stop points be determined? A basic principle is that the position should be liquidated at or before the point at which price movement causes a transition in the technical picture. For example, assume a trader decides to sell September natural gas after the mid-October downside breakout has remained intact for five days (see Figure 13.1). In this case, the protective buy stop should be placed no higher than the upper boundary of the July–October trading range, since the realization of such a price would totally transform the chart picture. Some of the technical reference points commonly used for placing protective stops include:

1. **Trend lines.** A sell stop can be placed below an uptrend line; a buy stop can be placed above a downtrend line. One advantage of this approach is that the penetration of a trend line will

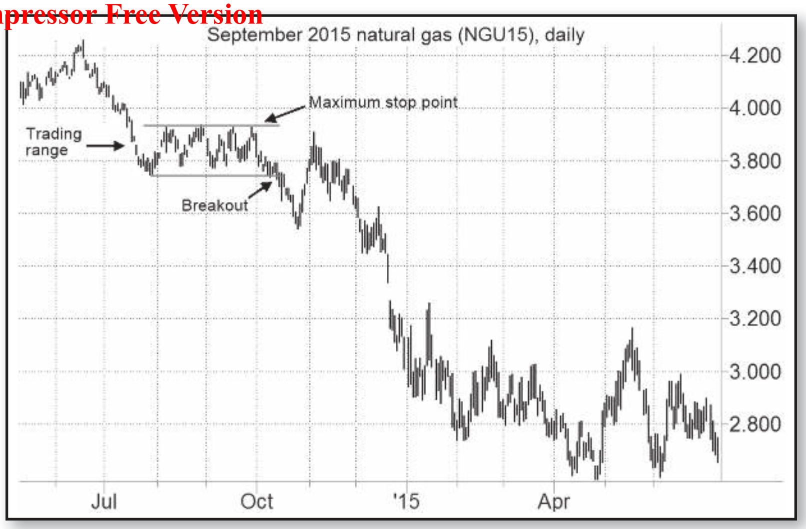

FIGURE 13.1 Stop Placement Following Trading Range Breakout: September 2015 Natural Gas

usually be one of the first technical signals in a trend reversal. Thus, this type of stop point will strongly limit the magnitude of the loss or the surrendered open profit. However, this attribute comes at a steep price: trend line penetrations are prone to false signals. As discussed in Chapter 6, it is common for trend lines to be redefined in the course of a bull or bear market.

2. **Trading range.** As illustrated in the preceding natural gas example, the opposite side of a trading range can be used as a stop point. Frequently, the stop can be placed closer (particularly in the case of broader trading ranges) because if the breakout is a valid signal, prices should not retreat too deeply into the range. Thus, the stop might be placed somewhere in the zone between the midpoint and the more distant boundary of the range. The near end of the trading range, however, would not be a meaningful stop point. In fact, retracements to this area are so common that many traders prefer to wait for such a reaction before initiating a position. (The advisability of this delayed entry strategy following breakouts is a matter of personal choice. In many instances it will provide better fills, but it will also cause the trader to miss some major moves.)
3. **Flags and pennants.** After a breakout in one direction of a flag or pennant formation, the return to the opposite end (or some point beyond) can be used as a signal of a price reversal, and by implication a point for placing stops. For example, in Figure 13.2 the downside penetration of a flag pattern in mid-August was quickly followed by a rebound above the same formation. This price action proved to be a precursor of a significant price advance.

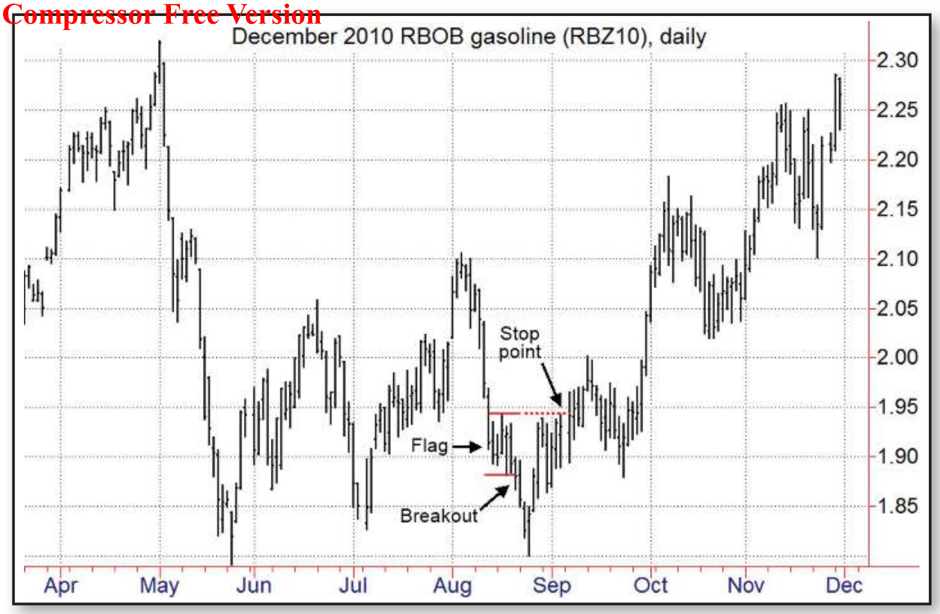

FIGURE 13.2 Stop Placement Following Flag Pattern Breakout: December 2010 RBOB Gasoline

4. **Wide-ranging days.** Similar to flags and pennants, after a breakout in one direction, the return to the opposite end can be used as a signal of a price reversal, and hence a point for placing stops. For example, in Figure 13.3 note how the return of prices back above the true high of the wide-ranging down day that formed in mid-March (after initially trading below this pattern) led to a strong rally.
5. **Relative highs and relative lows.** If the implied risk is not too great, the most recent relative high or relative low can be used as a stop point.[1] For example, assume a trader initiated a long position in December corn in response to the breakout above resistance in June (see Figure 13.4). In this case, the sell stop could be placed below either the May low or the June low.

Sometimes the risk implied by even the closest technically significant points may be excessive. In this case, the trader may decide to use a *money stop*—that is, a protective stop-loss point with no technical significance that is determined by the desired dollar risk level. For example, consider the plight of a trader in July 2008 who after the swift, steep (nearly $18/barrel) price break during the week ending July 18 is convinced the crude oil market has put in a major top (see Figure 13.5). The closest

[1] The specific definition of a relative low or relative high is somewhat arbitrary. (The following description is in terms of the relative low, but analogous commentary would apply to the relative high.) The general definition of a relative low is a day whose low is below the lows of the preceding and succeeding *N* days. The specific definition of a relative low will depend on the choice of *N*. A reasonable range for *N* is 5 to 15.

FIGURE 13.3 Stop Placement Following Wide-Ranging Day Breakout: June 2012 10-Year T-Note

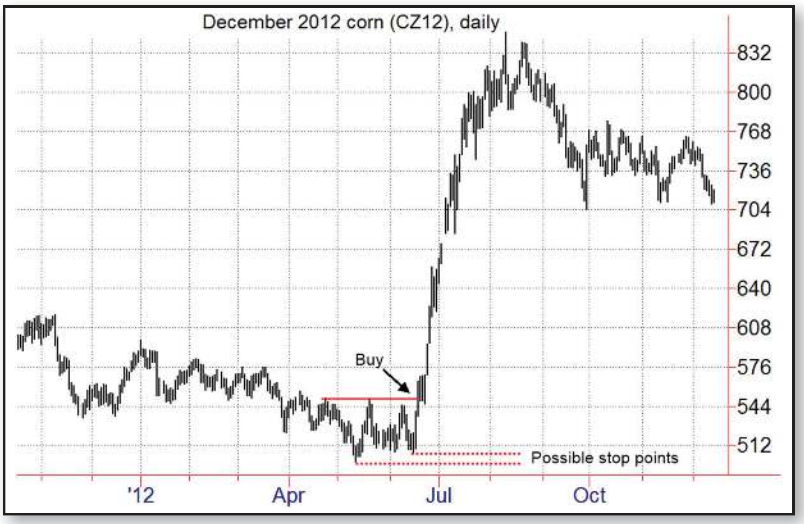

FIGURE 13.4 Stop Placement at Relative Lows: December 2012 Corn

FIGURE 13.5 Example of Market Where Money Stop Is Appropriate: December 2008 WTI Crude Oil

Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

meaningful stop point—the contract high (which is the nearest relative high)—would imply a risk of $17,850 per contract (assuming entry at the July 18 closing price)! Although risk can sometimes be reduced if the trader waits for a reaction before entering the market, such a retracement may not occur until the market moves substantially lower. Thus, in a situation in which the nearest meaningful stop point implies a very large risk, a market order accompanied by a money stop may represent the most viable trading approach.

Stops should be used not only to limit losses but also to protect profits. In the case of a long position, the stop should be raised intermittently as the market rises. Similarly, in a declining market, the stop should be lowered as the market declines. This type of stop is called a *trailing stop.*

Figure 13.6 illustrates the use of a trailing stop. Assume a trader implements a long position on the breakout above the upper boundary of the trading range, with a stop-loss liquidation plan keyed to relative lows. Specifically, the trader plans to liquidate the long position following a close below the most recent relative low with the reference point being revised each time the market moves to new high ground. (Of course, the stop condition may often be more restrictive. For example, the trader might require a specified number of closes below a previous low, or a minimum penetration of that low to activate the stop.) The initial stop-loss point would be a close below Stop 1, which is set at a level in the lower half of the trading range—a point that represents less risk than a stop at the more distant March 2009 relative low. Following the early June 2009 advance to new highs, the stop-loss reference point would be raised to the May low (Stop 2). Similarly, the stop reference points would be raised successively to the levels indicated by Stops 3 to 11. The position would have been stopped out on the decline below Stop 11 in March 2010.

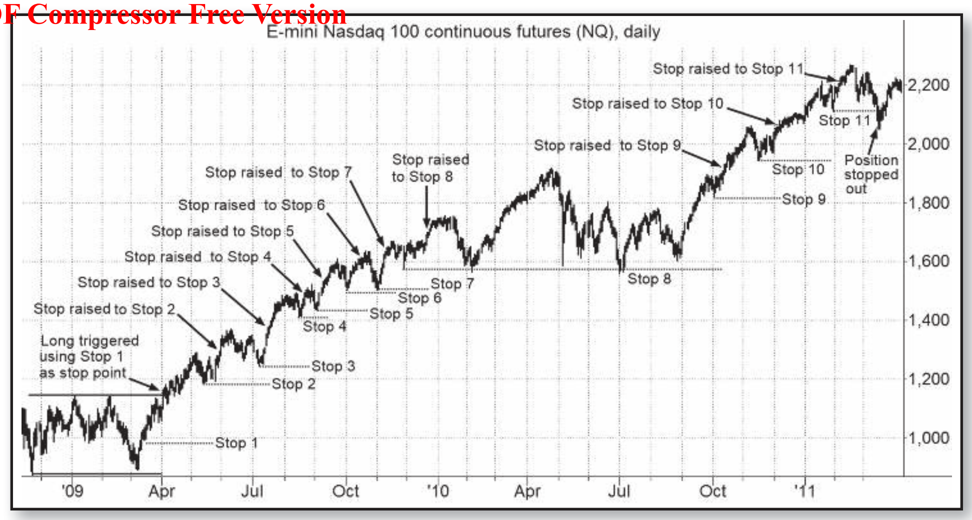

FIGURE 13.6 Trailing Stop: E-Mini Nasdaq 100 Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

As a general rule, stops should be changed only to reduce risk. Some traders who can't stand the thought of getting stopped out at the bottom of a move (top if short) may be diligent in placing a GTC stop order upon initiating the position, but then cancel the order when the market gets within range. This type of order has been derisively, albeit appropriately, referred to as a CIC (cancel if close) order. Revising the stop to allow greater risk defeats the entire purpose of the stop.

CHAPTER 14

Setting Objectives and Other Position Exit Criteria

It never was my thinking that made the big money for me. It was always my sitting. Got that? My sitting tight! It is no trick at all to be right on the market.

—Edwin Lefèvre

A trade is like the army—getting in is a lot easier than getting out. Provided the trader is adhering to money management principles, a losing trade presents little ambiguity; that is, liquidation would be indicated by a predetermined stop point. However, the profitable trade presents a problem (albeit a desirable one). How should the trader decide when to take profits? Myriad solutions have been proposed to this dilemma. The following sections explore some of the primary approaches.

Chart-Based Objectives

Many chart patterns are believed to provide clues regarding the magnitude of the potential price move. For example, conventional chart wisdom suggests that once prices penetrate the neckline of a head-and-shoulders formation, the ensuing price move will at least equal the distance from the top (or bottom) of the head to the neckline. As another example, many point-and-figure chartists claim that the number of columns that compose a trading range provides an indication of the potential number of boxes in a subsequent trend. (See discussion in Chapter 4 for an explanation of point-and-figure charting.) Generally speaking, chart patterns are probably considerably less reliable as indicators of price objectives than as trade signals.

Measured Move

This method is the essence of simplicity. The underlying premise is that markets will move in approximately equal-size price swings. Thus, if a market rallies 30 cents and then reacts, the implication is that the rally from the reaction low will approximate 30 cents. Although the measured move concept is so simple that it strains credibility, the approach offers reasonable guidelines more frequently than one might expect. When two or more of these objectives nearly coincide, it tends to enhance the reliability of the price area as an important objective zone.

Since price swings often span several contracts, it is useful to apply the measured move technique to longer-term price charts that link several contracts. Generally speaking, continuous futures charts are more appropriate than nearest futures charts for measured move analysis because, as was noted in Chapter 4 and further detailed in Chapter 5, continuous futures accurately reflect price swings, whereas nearest futures do not.

In Figure 14.1, the measured move objective that was fulfilled in July 2012 was the result of adding the amount of the December 2011–May 2012 rally (404.75¢) to the early June 2012 low of 667.25¢. Figure 14.2 shows two measured moves on a weekly chart. The first measured move target at 0.2711 (MM1), which was very close to the March 2015 relative low, was derived by subtracting the June–December 2014 decline of 0.0752 from the January 2015 high of 0.3462. The second measured move objective at 0.2297 (MM2), which was fairly close to the September 2015 low, was obtained by subtracting the January–March 2015 decline of 0.0818 from the late April high of

FIGURE 14.1 Measured Move: Soybean Continuous Futures

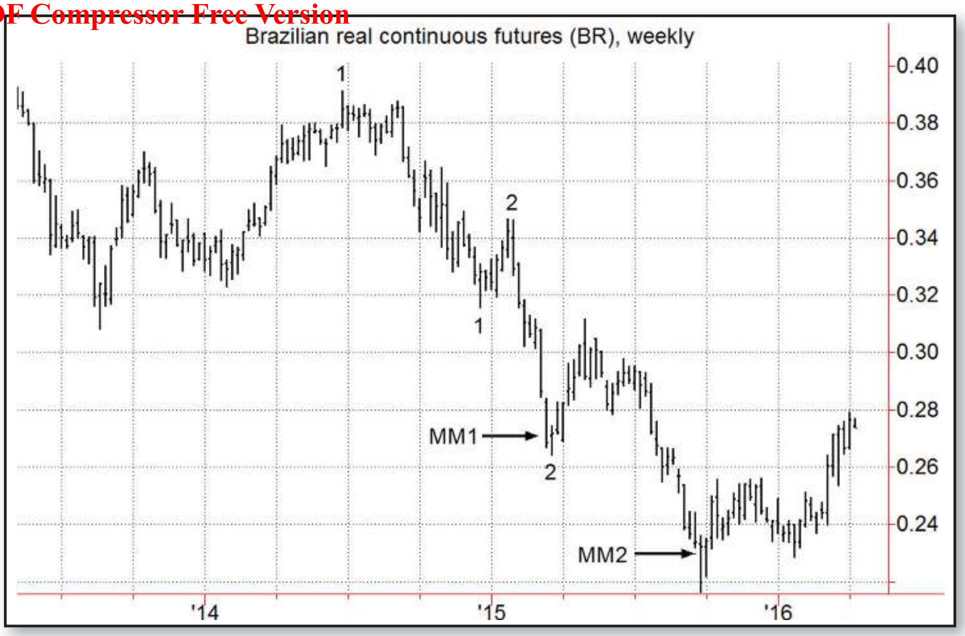

FIGURE 14.2 Measured Moves: Brazilian Real Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

0.3115. Figure 14.3 shows four measured move targets, three of which (MM1, MM3, and MM4) implied targets very near swing point highs.

Figure 14.4 illustrates a series of reasonably accurate measured move targets in frozen orange juice futures from mid-2012 to late 2013. Price didn't reach three of the targets (MM2, MM6, and MM9, represented by dashed lines), but missed only MM2 by a notable margin. Of the other six targets, all but MM4 represented quite advantageous exit points. Also, note that MM3 and MM5 signaled exit points at around the same level, reinforcing the target objective in that price vicinity.

Figure 14.5 provides another example of successive reasonably accurate measured move targets over a roughly two-year period. Note that the same price point can serve as the terminus of two different price swings (see October 2014 high with stacked 8 and 4), which can lead to two different measured move objectives based on that point (MM4 and MM8). This chart also provides an example of coincident measured move objectives: MM6, which is a projection based on the January–March 2014 upswing off the May low, occurred one tick away from MM8, which was the result of adding the June–October 2013 rally to the November low. MM4 and MM5 also signaled exits at approximately the same price level.

As Figures 14.4 and 14.5 illustrate, when there is more than one relevant price swing for deriving a measured move objective, there will be more than one measured move objective for the same projected low or high. When two or more of these objectives nearly coincide, it tends to enhance the reliability of the projected price area as an important target zone. Figure 14.6 provides a perfect example of two coinciding measured move price targets. The measured move objectives implied by the July 2014–March 2015 decline (MM5) and the May–October 2015 decline (MM6) coincided just above the actual market bottom formed in January 2016.

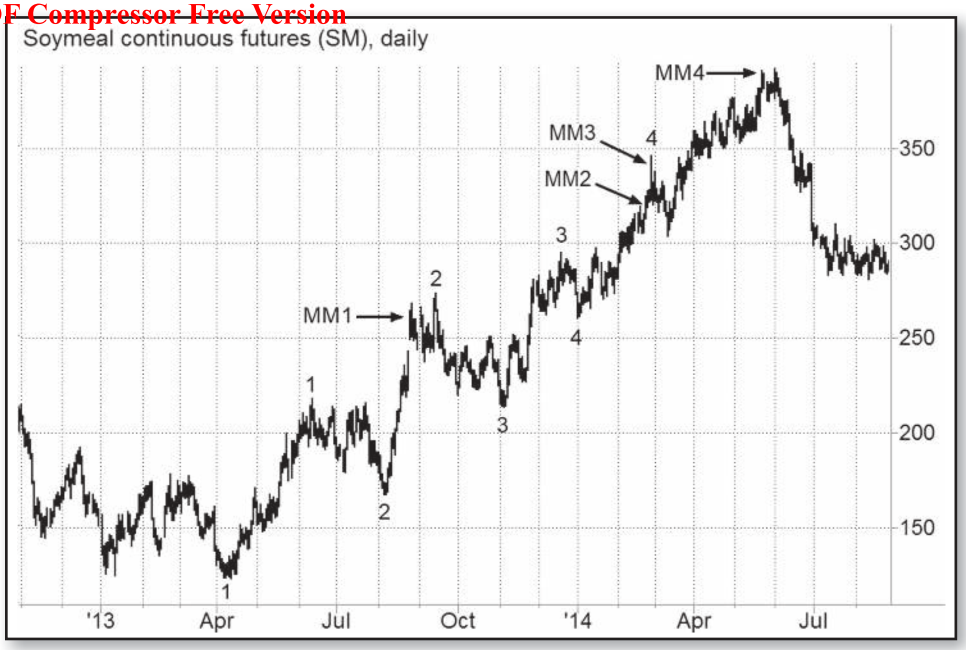

FIGURE 14.3 Measured Moves: Soymeal Continuous Futures

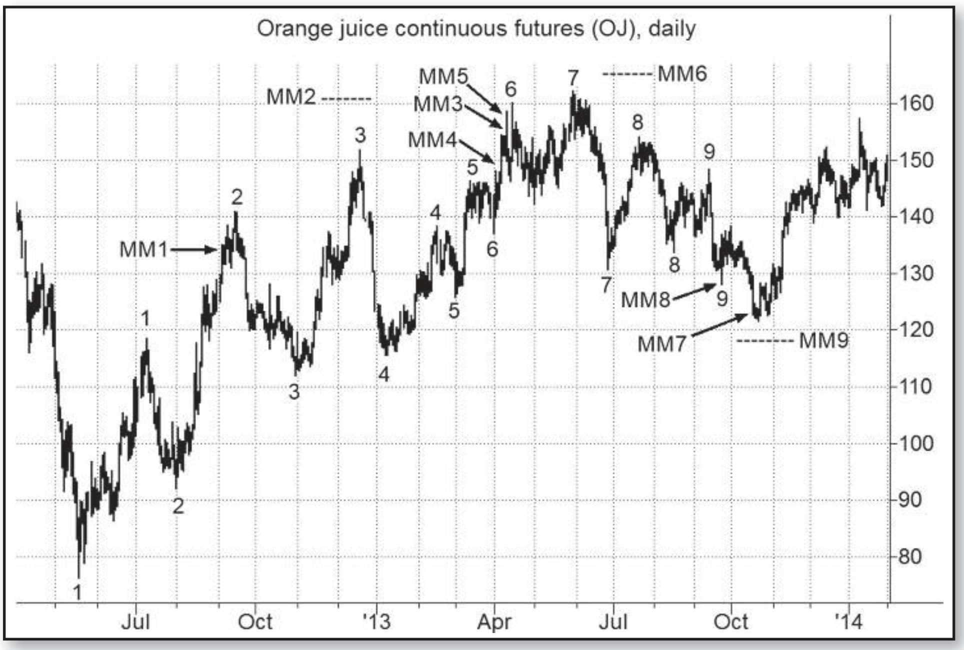

FIGURE 14.4 Measured Moves: Orange Juice Continuous Futures

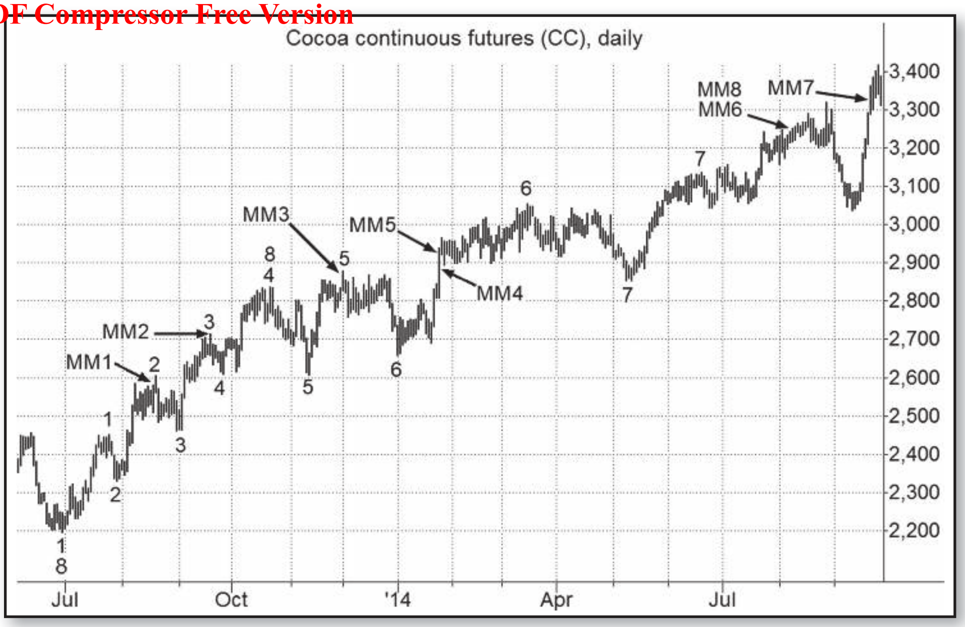

FIGURE 14.5 Concentration of Measured Move Targets: Cocoa Continuous Futures

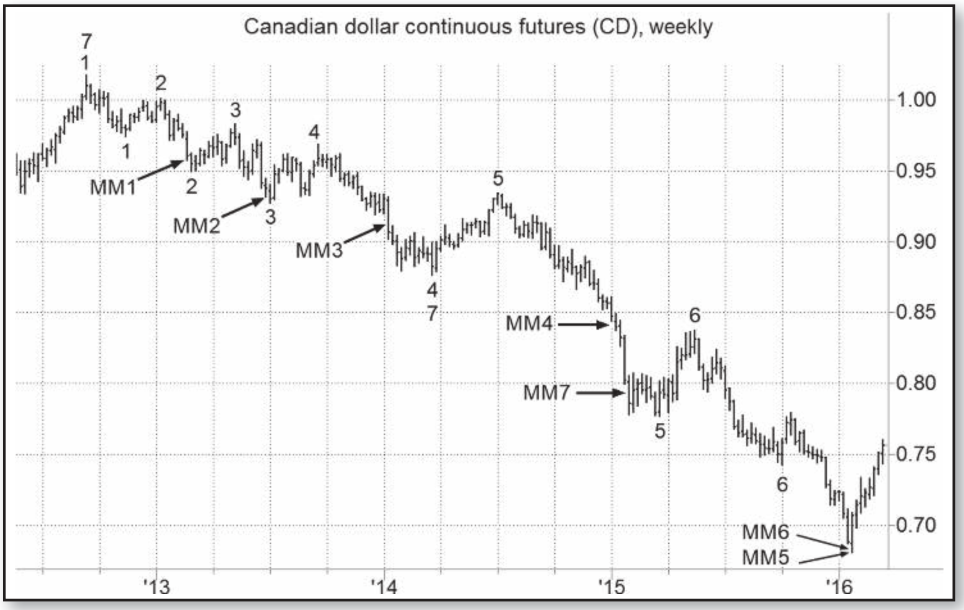

FIGURE 14.6 Concentration of Measured Move Targets: Canadian Dollar Continuous Futures

Rule of Seven

This method of setting objectives is an interesting and easy-to-use approach detailed in *Techniques of a Professional Commodity Chart Analyst* by Arthur Sklarew (Windsor Books, 1980). The rule of seven refers to a common set of multipliers used to determine objectives, which are derived by dividing 7 by 5, 4, 3, and 2, respectively. Thus, the multipliers are: $7 \div 5 = 1.4$, $7 \div 4 = 1.75$, $7 \div 3 = 2.33$, and $7 \div 2 = 3.5$. The products of each of these multipliers and the magnitude of the first price swing in a bull market are added to the low to obtain a set of price objectives. In a bear market, the products are subtracted from the high.

Sklarew suggests using the latter three multipliers (1.75, 2.33, and 3.5) for finding objectives in bull markets and the first three multipliers (1.4, 1.75, and 2.33) for deriving objectives in a bear market. In addition, he indicates objectives based on the lower multipliers are more meaningful if the reference price move (the price swing multiplied by the multipliers) is of extended duration (i.e., several months) and objectives based on the higher multipliers are more significant if a short-term price swing is used in the calculations. Of course, there will be some degree of subjectivity in this approach, since the perception of what constitutes the first price swing in a trend could vary from trader to trader.

The rule of seven is illustrated in Figure 14.7. (Note that this is the same chart that was used as Figure 14.3 to illustrate measured move objectives. Readers may find it instructive to compare the

FIGURE 14.7 Rule of Seven: Soymeal Continuous Futures

implications of these two approaches.) The first wave of the bull market that began in April 2013 was 94.30 points, measured from the April low to the June high. Following Sklarew's guidelines, because this is a bull market, we skip the first objective and use the second through fourth objectives, obtained using the multipliers 1.75, 2.33, and 3.5. The April 11 low, which is used to calculate all the objectives, was 123.90. The second objective is 288.90 [123.90 + (1.75 × 94.30)]. The third objective is 343.60 [123.90 + (2.33 × 94.30)]. The fourth objective is 454 [123.90 + (3.5 × 94.30)]. Note that objective 2 was just below the December 2013 relative high of 294.80, while objective 3 was just below the February 27 relative high of 346.10. The market failed to reach objective 4.

Figure 14.8 (which repeats Figure 14.6) illustrates the rule of seven for an extended bear market in Canadian dollar continuous futures. The chart intentionally shows two sets of objectives based on using different lows (A and B) to define the initial leg of the downtrend. In both cases, the September 2012 high was used as the initial high reference price. The first wave of this bear market using low A in March 2013 was 0.0674 points, while using low B in March 2014 the first wave was 0.1407 points. Following Sklarew's guidelines, since this is a bear market, we use the first through third objectives (obtained using the multipliers 1.4, 1.75, and 2.33). The products of these three multipliers and the two initial price swings are subtracted from the high of the move to obtain the two sets of downside objectives. Of the three objectives that referenced low A, only Objective 2 was

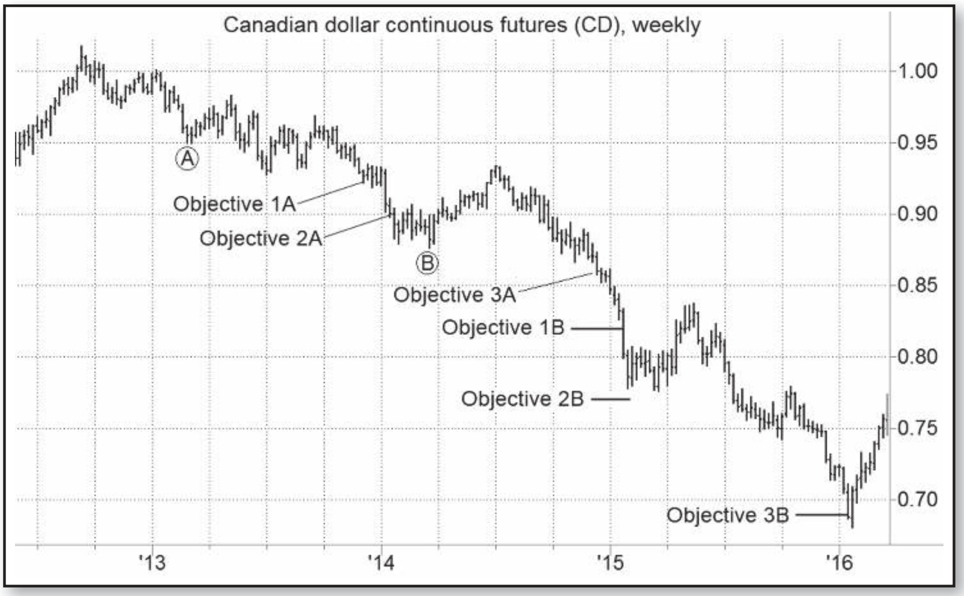

FIGURE 14.8 Rule of Seven: Canadian Dollar Continuous Futures

fairly close to a relative low (during the January–March 2014 consolidation). Among the objectives using low B, Objective 2 was just below the March 2015 relative low, while Objective 3 was just above the January 2016 low.

Support and Resistance Levels

Points near support levels provide a reasonable choice for setting initial objectives on short positions. For example, the indicated objective zone in Figure 14.9 is based on support anticipated in the area of two prior relative lows. Similarly, prices near resistance levels can be used for setting initial objectives on long positions. For example, the indicated objective in Figure 14.10 is based on resistance implied by the two previous highs in late 2009 and early 2010. In Figure 14.11, an upside objective for British pound prices after the early 2009 bottom was implied by the late 2005 relative low, a level that continued to function as a ceiling for prices over the next several years (a case of former support becoming resistance, as discussed in Chapter 8).

Generally speaking, support and resistance levels usually represent only temporary rather than major objectives. Consequently, in using this approach, it is advisable to seek to reenter the position at a better price if a reaction does develop.

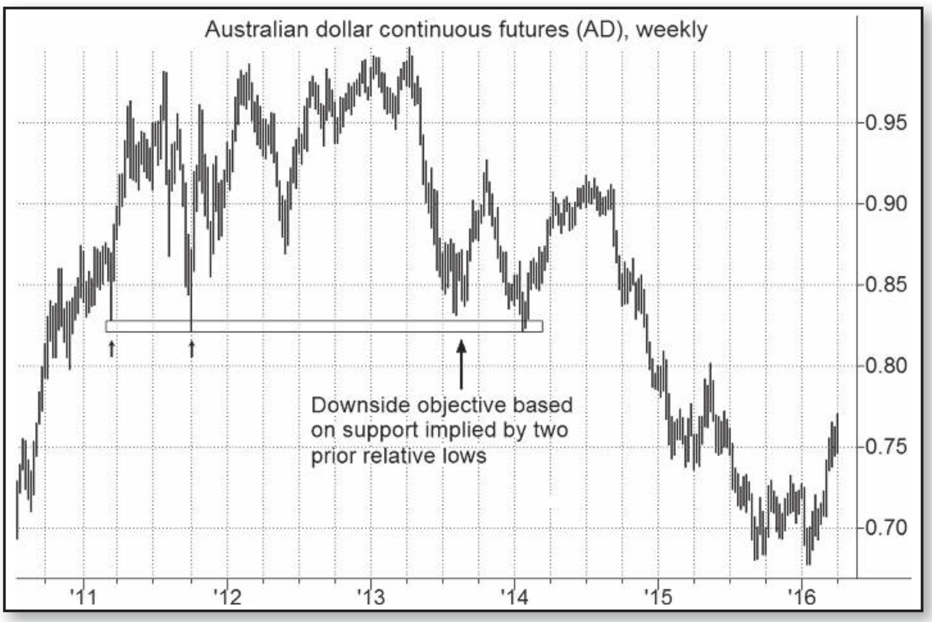

FIGURE 14.9 Downside Objective at Support Zone: Australian Dollar Continuous Futures

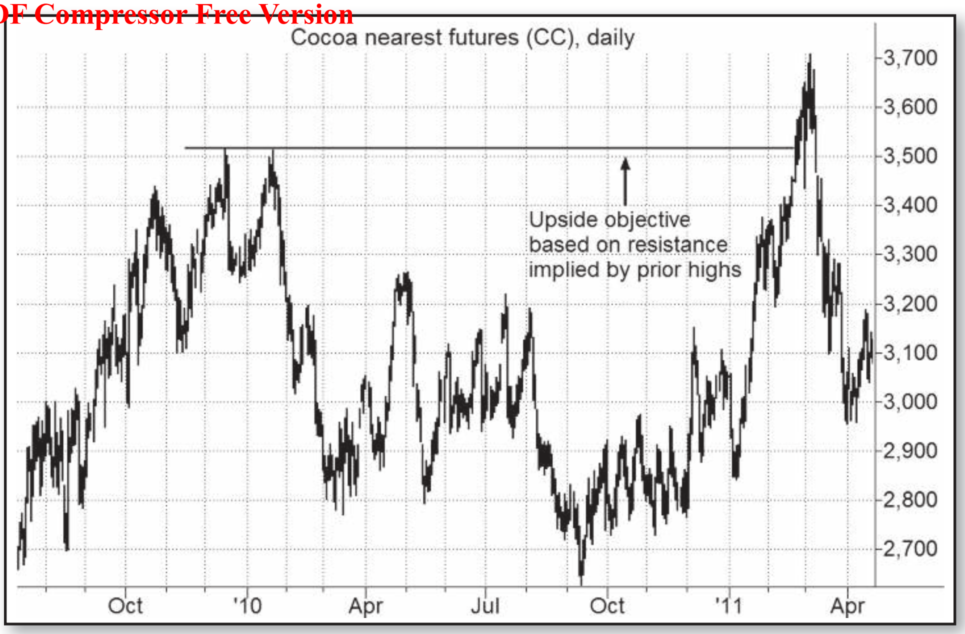

FIGURE 14.10 Upside Objective at Resistance Level: Cocoa Nearest Futures

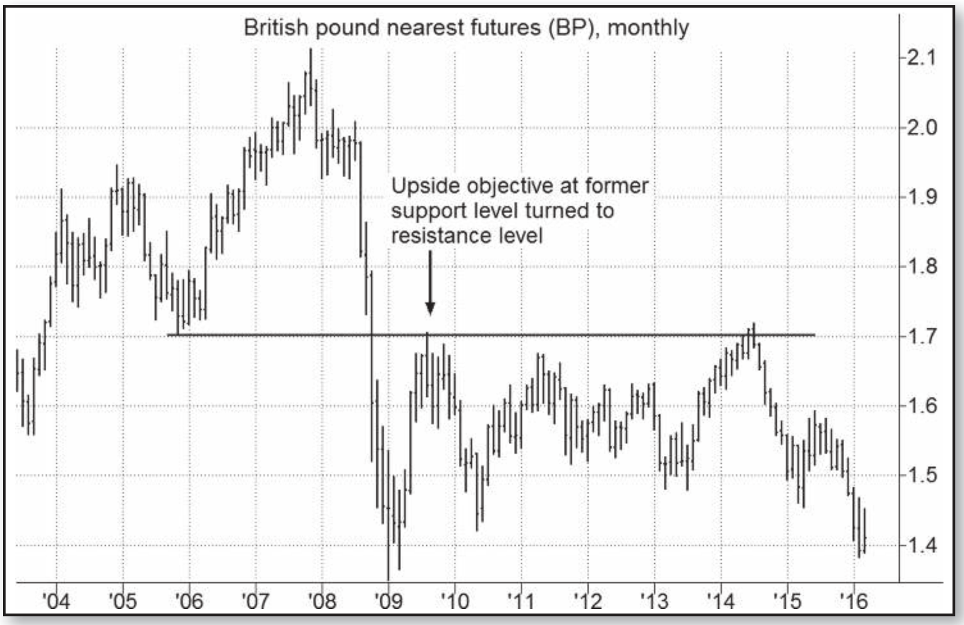

FIGURE 14.11 Upside Objective at Former Support Turned Resistance: British Pound Nearest Futures

Overbought/Oversold Indicators

Overbought/oversold indicators are technical measures intended to reflect when prices have risen or fallen too sharply and are thus vulnerable to a reaction. Figure 14.12 illustrates the relative strength index (RSI), which provides an example of an overbought/oversold indicator.[1] The RSI has a range of values between 0 and 100. Based on the standard interpretation, levels above 70 suggest an overbought condition, while levels below 30 suggest an oversold condition.

The choice of specific overbought/oversold boundaries is a subjective one. For example, instead of 70 and 30, one might use 75 and 25, or 80 and 20. The more extreme the selected threshold levels, the closer the overbought/oversold signals will be to market turning points, but the greater the number of such points that will be missed.

The buy (up) arrows in Figure 14.12 denote points at which the RSI crosses below 30—that is, reaches an oversold condition that can be viewed as a signal to liquidate short positions. The sell (down) arrows denote points at which the RSI crosses above 70—that is, reaches an overbought condition that can be viewed as a signal to liquidate long positions.

Although the overbought/oversold signals in Figure 14.12 provide some reasonably good position liquidation signals in the latter half of the chart (mid-April 2015 forward), the signals before that

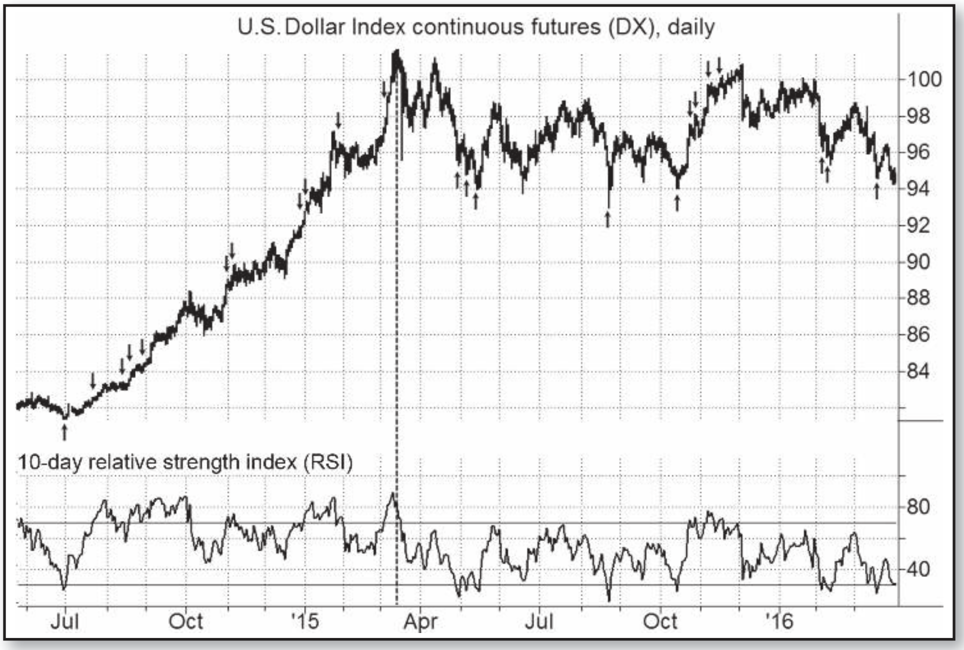

FIGURE 14.12 Relative Strength Index in Trend and Trading Range Conditions: U.S. Dollar Index Continuous Futures

[1] The RSI was originally introduced in J. Welles Wilder, Jr., *New Concepts in Technical Trading Systems* (Winston-Salem, NC: Hunter Publishing, 1978).

point—when the market was in a strong uptrend—were almost all terrible. The 27 percent rally off the July 2014 low that ultimately extended into March 2015 generated 10 overbought signals, four of which occurred in rapid succession in the first two months of the rally. Only the final two signals during this period, in late January and early March 2015, could be considered relatively timely. This example hints at both the benefits and drawbacks of using overbought/oversold indicators as liquidation signals. The approach will usually work well when the market is in a trading range, but will fail miserably during strong trending phases.

The derivation and interpretation of various technical indicators are discussed in detail in Chapter 11.

DeMark Sequential

As discussed in Chapter 11, all the popular overbought/oversold indicators (e.g., RSI, moving average convergence-divergence [MACD], stochastic) are very highly correlated with each other. Tom DeMark's sequential, which is intended to signal points where the market is fully extended and vulnerable to a major trend reversal, represents a completely different and original overbought/oversold indicator. The sequential methodology falls within the domain of pattern recognition. The sequential is fully described in a 48-page chapter in Tom DeMark's book *The New Science of Technical Analysis* (John Wiley & Sons, 1994). The following brief summary of the technique is intended to give a general sense of the approach. Readers interested in a fully detailed explanation of the sequential, which includes several additional qualifying conditions and a discussion of various alternative trade entry and exit rules, are referred to DeMark's text.

The fulfillment of the sequential *buy* condition involves three basic stages:

1. **Setup.** The setup requires nine or more consecutive closes that are lower than the corresponding closes four trading days earlier.
2. **Intersection.** This condition requires that the high of any day on or after the eighth day of the setup exceed the low of any day three or more trading days earlier. Essentially, this is a minimal qualifying condition that ensures that the buy setup will not be deemed complete in a "waterfall" price slide.
3. **Countdown.** The countdown stage begins once the previous two conditions have been fulfilled. Starting from 0, the countdown increases by one on each day with a close lower than the low two days earlier. A sequential buy signal is generated once the countdown reaches 13. In contrast to the setup stage, countdown days do not need to be consecutive. The countdown is canceled if any of the following three conditions arise:
    a. There is a close that exceeds the highest intraday high during the setup stage.
    b. A *sell* setup occurs (i.e., nine consecutive closes above the corresponding closes four days earlier).
    c. Another buy setup occurs before the buy countdown is complete. In this situation, the new buy setup takes precedence, and the countdown restarts from 0 once the intersection condition is met.

The fulfillment of the sequential *sell* conditions are analogous:

1. **Setup.** The setup requires nine or more consecutive closes that are higher than the corresponding closes four trading day earlier.
2. **Intersection.** This condition requires that the low of any day on or after the eighth day of the setup is lower than the high of any day three or more trading days earlier. Essentially, this is a minimal qualifying condition that ensures that the sell setup will not be deemed complete in a "runaway" rally.
3. **Countdown.** The countdown stage begins once the previous two conditions have been fulfilled. Starting from 0, the countdown increases by one on each day with a close higher than the high two days earlier. A sequential sell signal is generated once the countdown reaches 13. In contrast to the setup stage, countdown days do not need to be consecutive. The countdown is canceled if any of the following three conditions arise:
    a. There is a close that is below the lowest intraday low during the setup stage.
    b. A *buy* setup occurs (i.e., nine consecutive closes below the corresponding closes four days earlier).
    c. Another sell setup occurs before the sell countdown is complete. In this situation, the new sell setup takes precedence, and the countdown restarts from 0 once the intersection condition is met.

Figures 14.13 through 14.17 provide illustrations of markets that fulfilled the complete sequential process. In each case, the setup, intersection, and countdown stages are marked on the charts; the final bar of the setup stage is highlighted with a boldfaced **9,** while the final bar of the countdown phase is marked with a boldfaced **13.** The preceding description will be clearer if read in conjunction with an examination of these charts.

Figure 14.13 provides an illustration of a sequential sell signal in June 2016 10-year T-note futures. Note that in this case, the first day of the countdown stage (which occurred three days after the end of the setup stage) also fulfilled the intersection requirement (a bar with a low below the high of a day three or more days earlier). The countdown phase completed on February 11, the day that marked the highest high and close of the upmove. Figure 14.14, which shows the June 2016 gold contract, provides an example of a sequential buy. As was the case in Figure 14.13, the first day of the countdown stage also marked the fulfillment of the intersection requirement. The completion of the countdown stage coincided with the mid-December 2015 low.

Figure 14.15 provides another example of a sequential buy, this time in the May 2016 soybean contract. In this case the intersection requirement occurred on the eighth bar of the setup phase, while the countdown phase didn't begin until nine days after the end of the setup phase. The countdown completed in early March 2016, the day with the lowest low of the move and one day after the lowest close. (*Note:* Figures 14.14 and 14.15 reflect day-session-only data.)

The sequential rules can also be applied to bar charts for time periods other than daily. Figure 14.16 illustrates a sequential sell on a monthly copper continuous futures chart. Here, the end of the setup stage, the beginning of the countdown stage, and the fulfillment of the intersection requirement all occur on the same bar (month). The market peaked at month 11 of the countdown phase, but the real

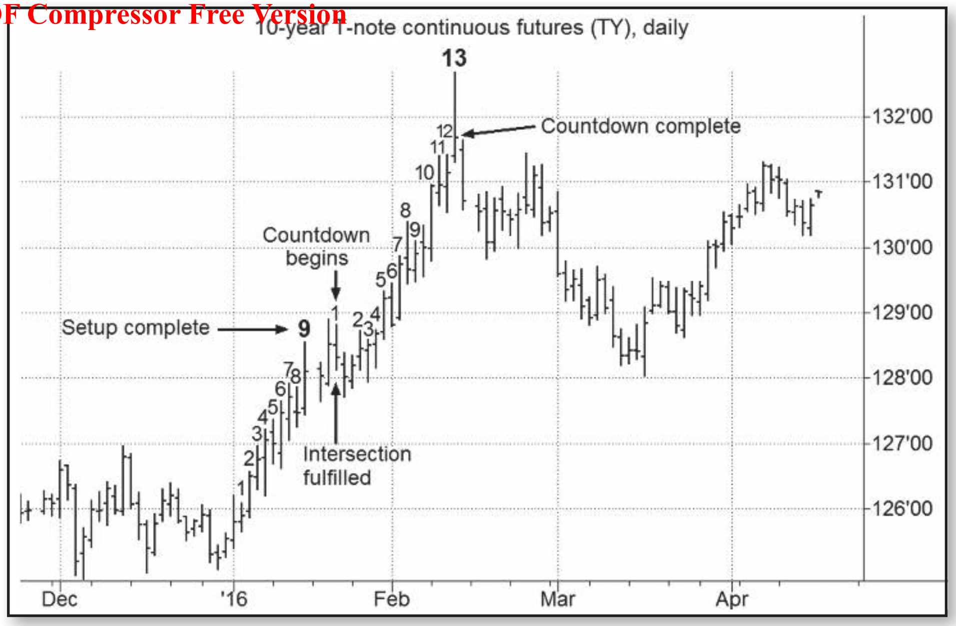

FIGURE 14.13 DeMark Sequential: June 2016 10-Year T-Note Continuous Futures
Source for sequential signals: DeMark Analytics (www.demark.com)

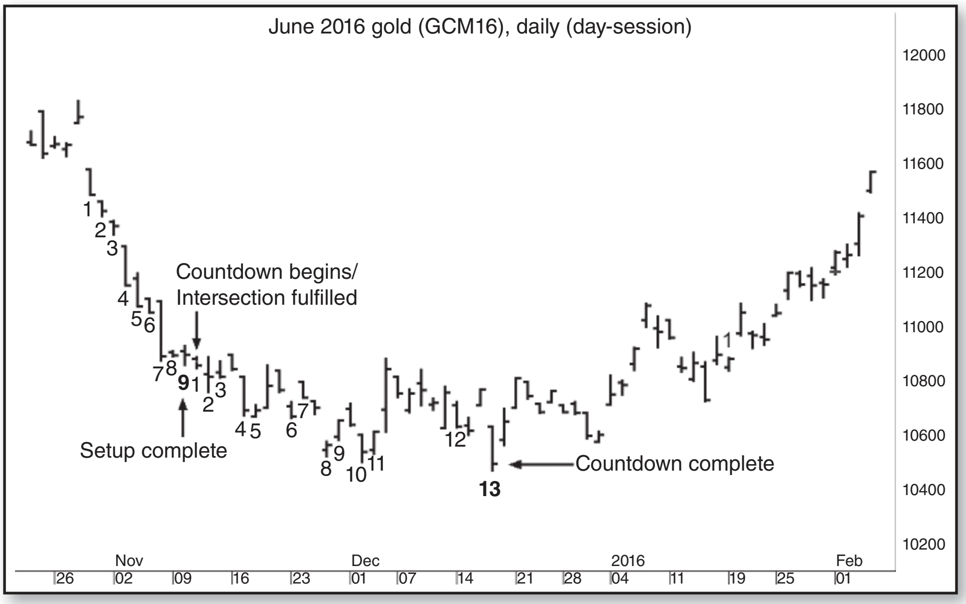

FIGURE 14.14 DeMark Sequential: June 2016 Gold
Source for sequential signals: Copyright 2016, DeMark for CQG, www.demark.com

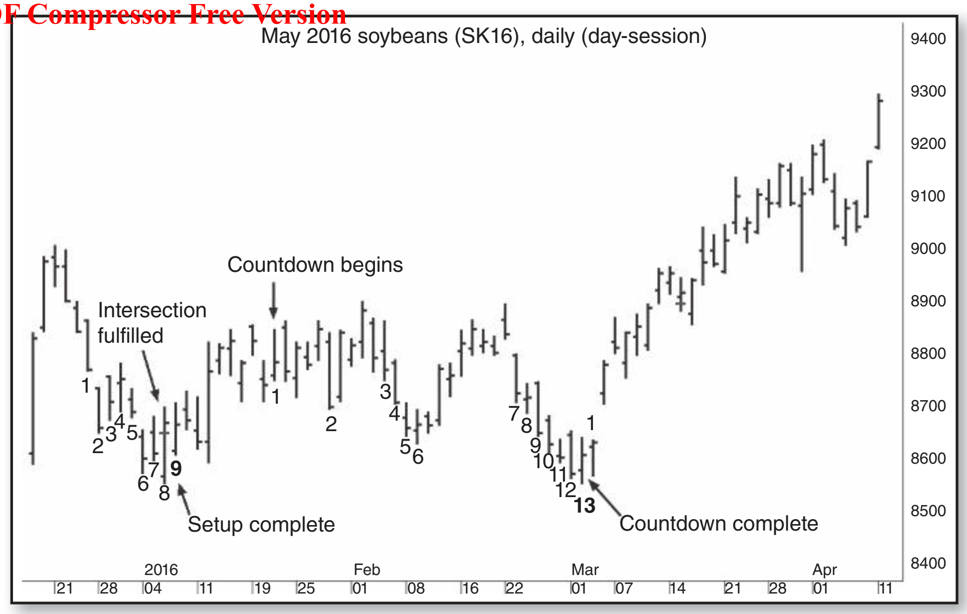

FIGURE 14.15 DeMark Sequential: May 2016 Soybeans

Source for sequential signals: Copyright 2016, DeMark for CQG, www.demark.com

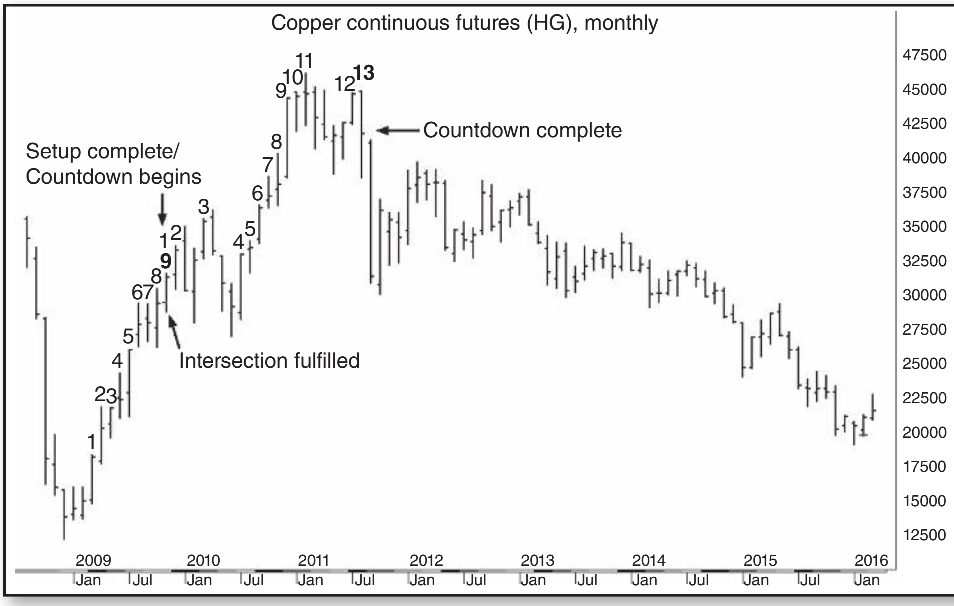

FIGURE 14.16 DeMark Sequential: Copper Continuous Futures

Source for sequential signals: Copyright 2016, DeMark for CQG, www.demark.com

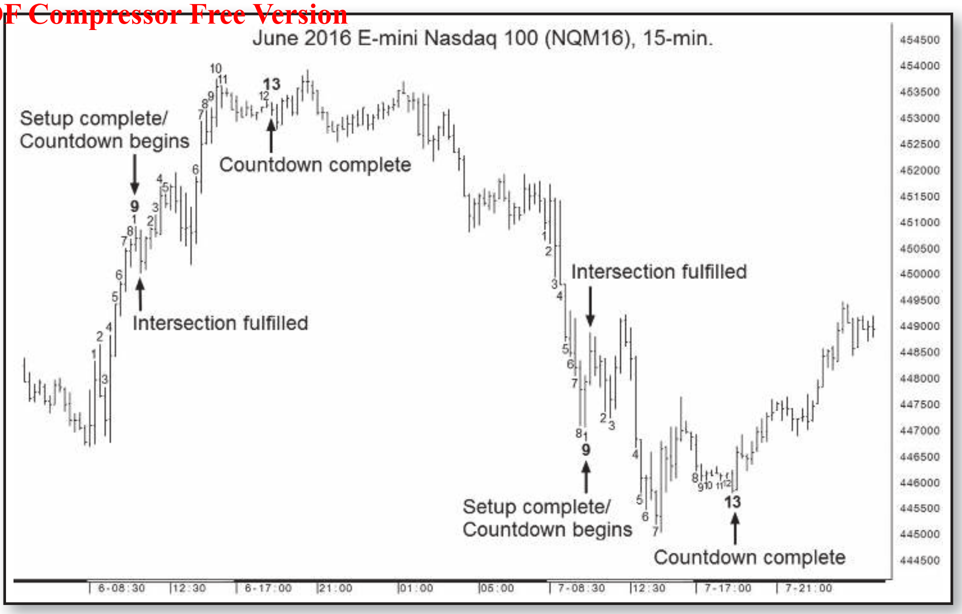

FIGURE 14.17 DeMark Sequential: June 2016 E-Mini Nasdaq 100
Source for sequential signals: Copyright 2016, DeMark for CQG, www.demark.com
Chart from CQG, Inc. © 2017 All rights reserved worldwide. Signal from Demark Analytics.

reversal did not occur until after the completion of countdown six months later. Figure 14.17 shows completed sequential sell and buy setups on an intraday chart (15-minute bars). The sell setup completed during a consolidation near the top of the rally, while the buy setup completed a bit above the low of the subsequent decline, but right at the start of the first extended rally after the low.

The preceding examples were obviously selected with hindsight to illustrate the methodology. Of course, in real-life trading, the accuracy of the DeMark sequential approach will not approach the uniformly near-perfect signals provided by the previous set of examples. If it did, all anyone would need to do would be to trade all sequential signals and retire a multimillionaire. Nevertheless, these examples should demonstrate that sequential can be a very powerful tool, with the capability of providing extraordinary timing signals. Sequential also has the advantage of being inversely correlated to trend-following approaches that typically dominate the technical tool bag. For these reasons, many traders might find DeMark's sequential a very useful addition to their overall trading methodology.

Contrary Opinion

The theory of contrary opinion suggests that whenever a large majority of speculators are bullish, those who want to be long are already long. Consequently, there will be a paucity of potential new buyers, and the market will be vulnerable to a downside reaction. An analogous interpretation would apply when the majority of traders are bearish. Contrary opinion measures are based on either surveys of market advisory recommendations or surveys of traders and implicitly assume these opinions represent a reasonable proxy for overall market sentiment. The overbought and oversold thresholds in contrary opinion indexes will vary with the source.

Although contrary opinion is undoubtedly a sound theoretical concept, the Achilles' heel of this approach is the difficulty of measuring market sentiment accurately. Contrary opinion measures provided by existing services have frequently signaled major turning points. On the other hand, it is also not unusual for a contrary opinion index to stay high while the market continues to climb, or to stay low as the market continues to slide. On balance, this method provides useful information as long as it is not used as the sole trading guideline.

Trailing Stops

The use of trailing stops may be among the least glamorous, but most sensible, methods of determining a trade exit point. Although one will never sell the high or buy the low using this method, the approach comes closest to the ideal of permitting a profitable trade to run its course. Trailing stops were detailed in Chapter 13.

Change of Market Opinion

This method of exiting trades represents another approach with very little flash, but lots of common sense. In this case, the trader sets no predetermined objectives at all, but rather maintains the position until her market opinion changes to at least neutral.

CHAPTER 15

The Most Important Rule in Chart Analysis

The market is like a flu virus—as soon as you think you have it pegged, it mutates into something else.

—Wayne H. Wagner

Failed Signals

A failed signal is among the most reliable of all chart signals. When a market fails to follow through in the direction of a chart signal, it very strongly suggests the possibility of a significant move in the opposite direction. For example, in Figure 15.1 note how the market abruptly reversed course after breaking out above the high of the July–August 2013 consolidation in WTI crude oil. If the upside penetration signal were valid, the market should not have retreated back to the lower portion of the consolidation and certainly not below its lower boundary. The fact that such a retracement occurs almost immediately following the breakout strongly suggests a "bull trap." Such price action is consistent with the market's rising just enough to activate stop orders lying beyond the boundary of the range, but uncovering no additional buying support after the breakout—an indication of a very weak underlying technical picture. In effect, the immediate failure of the apparent buy signal can be viewed as a strong indication the market should be sold.

Now that we have established the critical importance of failed signals, the following sections detail various types of failed signals, along with guidelines as to their interpretation and trading implications.

Bull and Bear Traps

Bull and bear traps are major breakouts that are soon followed by abrupt, sharp price reversals, in stark contrast to the price follow-through that is expected to follow breakouts. In my experience, this type of counter-to-anticipated price action is among the most reliable indicators of major tops and bottoms.

FIGURE 15.1 Bull Trap: WTI Crude Oil Continuous Futures

An example of a bull trap was provided in the previous section (Figure 15.1). Another instance of a bull trap was the June 2015 peak in RBOB gasoline (see Figure 15.2). After rallying from January to May 2015, the market consolidated for roughly one month before breaking out to new highs in mid-June. However, the market quickly reversed back into the trading range, and by mid-July prices had broken below the range's lower boundary, setting the stage for a multimonth downtrend.

Analogous to the bull trap, in the case of a bear trap, the market falls just enough to trigger resting stops below the low end of a trading range, but fails to uncover any additional selling pressure after the breakout—an indication of substantial underlying strength. In effect, the immediate failure of a sell signal can be viewed as a signal the market should be bought.

Figure 15.3 shows a bear trap that marked the 2014 low in U.S. Dollar Index futures. In May the market broke below the lower boundary of a long-standing trading range but reversed two days later to close back above that threshold. This price action proved to be the beginning of the market's largest rally in more than a decade.

Figure 15.4 provides another example of a bear trap. Corn prices, which had been trending lower since late summer 2012, entered a trading range in November–December 2013. The market broke below the downside of the range in early January 2014, falling more than 2 percent over the next two days before reversing sharply and returning to the midpoint of the range. May corn futures subsequently surged approximately 25 percent over the next three months.

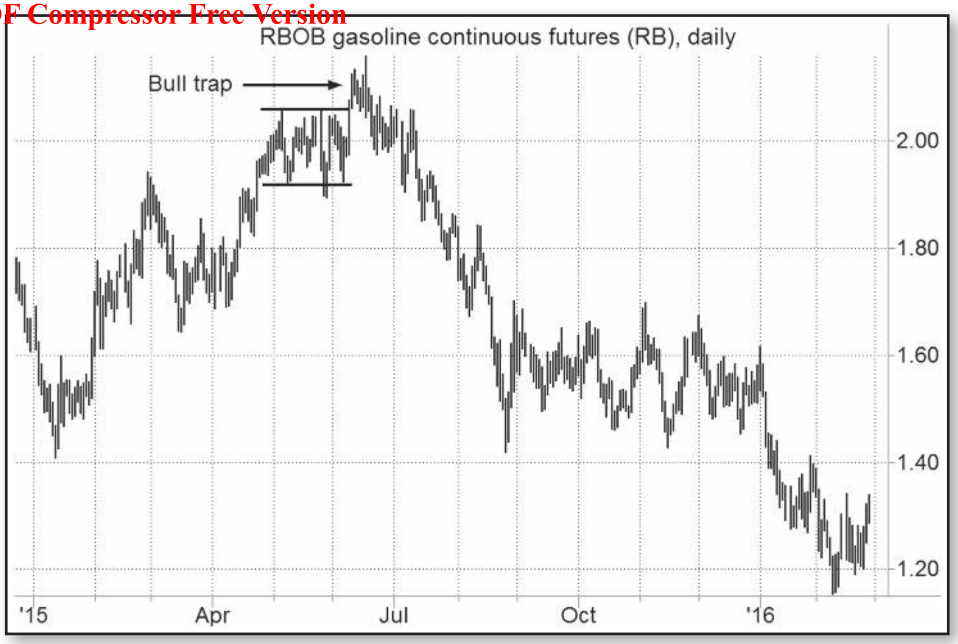

FIGURE 15.2 Bull Trap: RBOB Gasoline Futures

FIGURE 15.3 Bear Trap: U.S. Dollar Index Futures

FIGURE 15.4 Bear Trap: May 2014 Corn Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

How much of a pullback is required to indicate a bull or bear trap has occurred? The following are several possible confirmation conditions:

Initial price confirmation. A price retracement to the midpoint of the consolidation that preceded the breakout.

Strong price confirmation. A price retracement to the more distant boundary (lower for bull trap; upper for bear trap) of the consolidation that preceded the breakout.

Time confirmation. The failure of the market to return to the extreme price witnessed following the breakout within a specified amount of time (e.g., four weeks).

The trade-off between initial and strong price confirmations is that the former will provide better entry levels in trading bull and bear traps, whereas the latter will provide more reliable signals. The time confirmation condition can be used on its own or in conjunction with the two price confirmation conditions. Figures 15.5 through 15.8 repeat Figures 15.1 through 15.4, adding each of the three confirmation conditions (using four weeks for the time confirmation condition). Note the time confirmation can occur before both price confirmation conditions, after both price confirmation conditions (as is the case in Figures 15.6 and 15.8), or between the price confirmations (Figures 15.5 and 15.7).

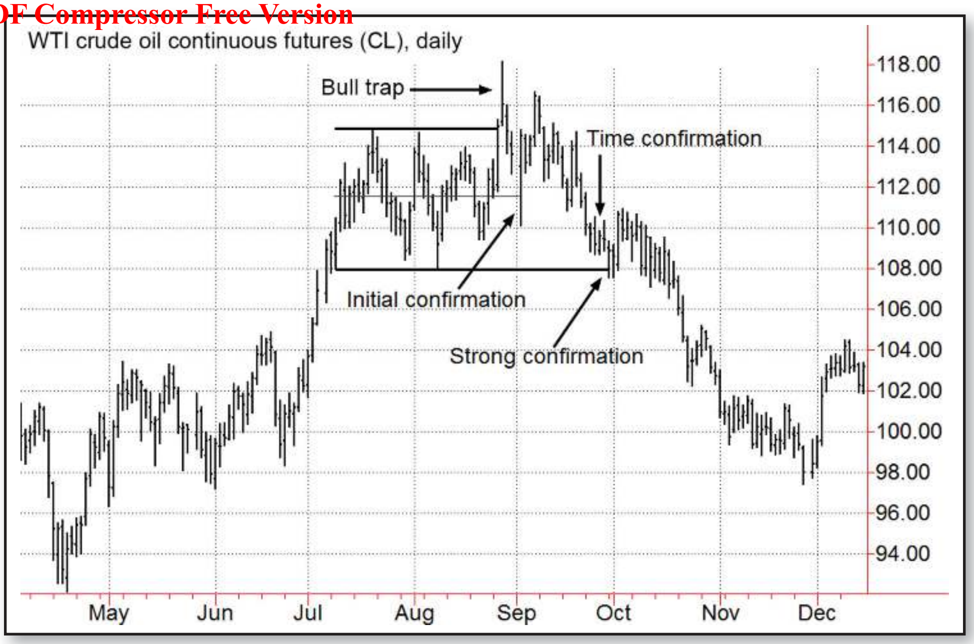

FIGURE 15.5 Bull Trap Confirmation Conditions: WTI Crude Oil Continuous Futures

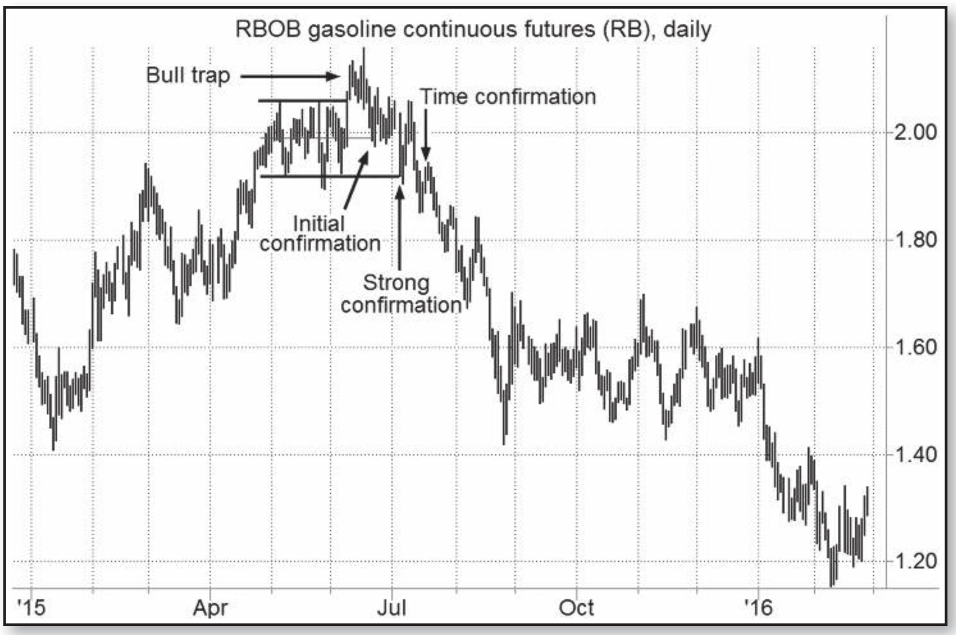

FIGURE 15.6 Bull Trap Confirmation Conditions: RBOB Gasoline Continuous Futures

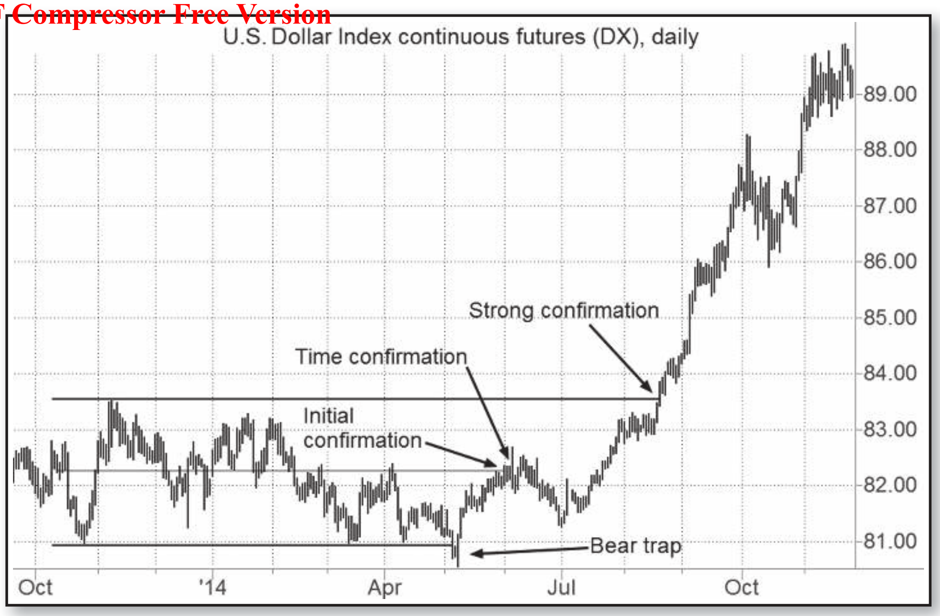

FIGURE 15.7 Bear Trap Confirmation Conditions: U.S. Dollar Index Continuous Futures

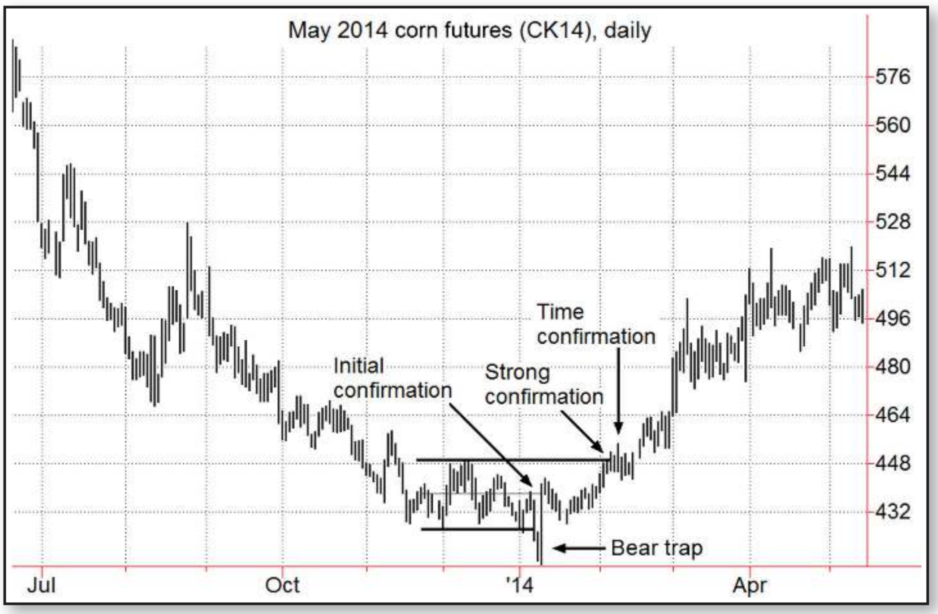

FIGURE 15.8 Bear Trap Confirmation Conditions: May 2014 Corn Futures

A bull trap signal would be invalidated if the market returned to the breakout high. Similarly, a bear trap signal would be invalidated if the market returned to the breakout low. More sensitive conditions could be used to invalidate bull or bear trap signals once the market has moved sufficiently in the direction of the signal or a specified amount of time has elapsed. An example of such a condition would be the return of prices to the opposite boundary of a consolidation once a strong price confirmation signal was received (e.g., in the case of a bull trap, a return to the top of the consolidation after prices broke to below the low end of the consolidation). An example of a more sensitive combined price/time invalidation signal would be the return of prices to the median of a consolidation (i.e., the initial *price* confirmation point for bull and bear trap signals) at any time four or more weeks after a strong price confirmation was received. The more sensitive the selected invalidation condition, the smaller the loss on an incorrect call of a bull or bear trap, but the greater the chance that a correct trade will be abandoned prematurely.

If the selected invalidation condition does not occur, a trade implemented on a bull or bear trap signal would be held until a price objective or other trade liquidation condition was met or until there was evidence of an opposite direction trend reversal.

False Trend Line Breakouts

As discussed in Chapter 6, trend lines are particularly prone to false breakouts. Such false breakouts can be used as signals for trading in the direction opposite to the breakout. In fact, in my opinion, false trend line breakout signals are considerably more reliable than conventional trend line breakout signals. In the case of a downtrend, a false trend line breakout would be confirmed if the market closed below the trend line a specified number of times (e.g., two, three) following an upside breakout. Similarly, in the case of an uptrend, a false trend line breakout would be confirmed if the market closed above the trend line a specified number of times following a downside breakout.

Figure 15.9 provides an example of a false breakout of an uptrend line in 10-year T-note futures. The September downside breakout of the uptrend line was soon followed by a break above the line. The indicated failure signal is based on an assumed requirement of two closes above the line for confirmation. Figure 15.10 provides a similar example in the E-mini Nasdaq 100 futures.

It is quite possible for a chart to yield multiple successive false trend breakout signals in the process of a trend line being redefined. In Figure 15.11 the initial upside penetration of the prevailing downtrend line occurred in mid-March. Prices quickly retreated back below the line, with the indicated failure signal assumed to be triggered by the second close below the line. Another false breakout occurred about a month later based on the redefined trend line using the March relative high. Prices retreated below this downtrend line several days later, yielding another false trend breakout signal.

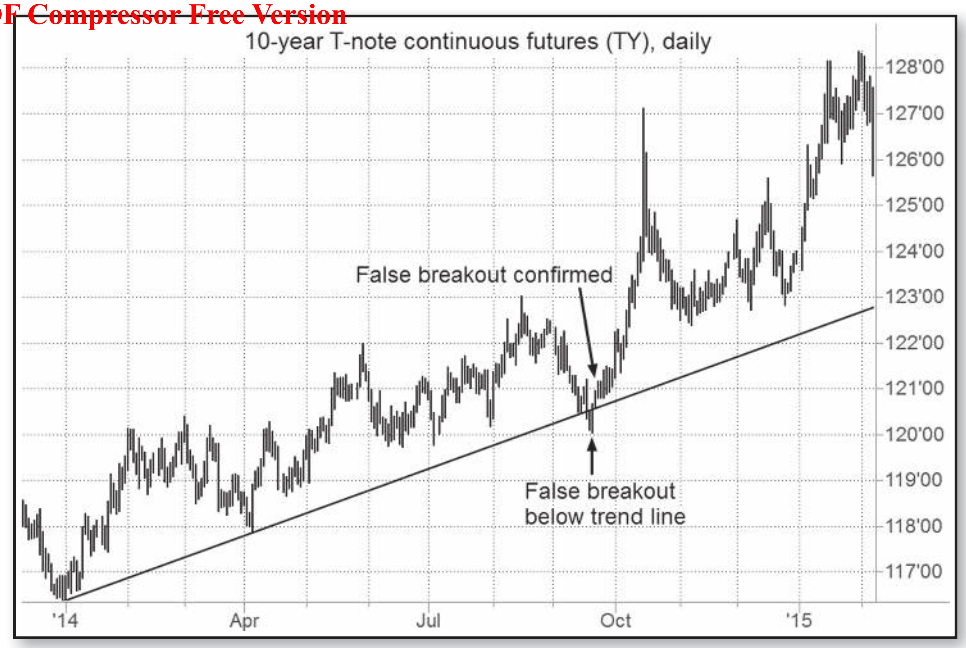

FIGURE 15.9 False Breakout of Uptrend Line: 10-Year T-Note Continuous Futures

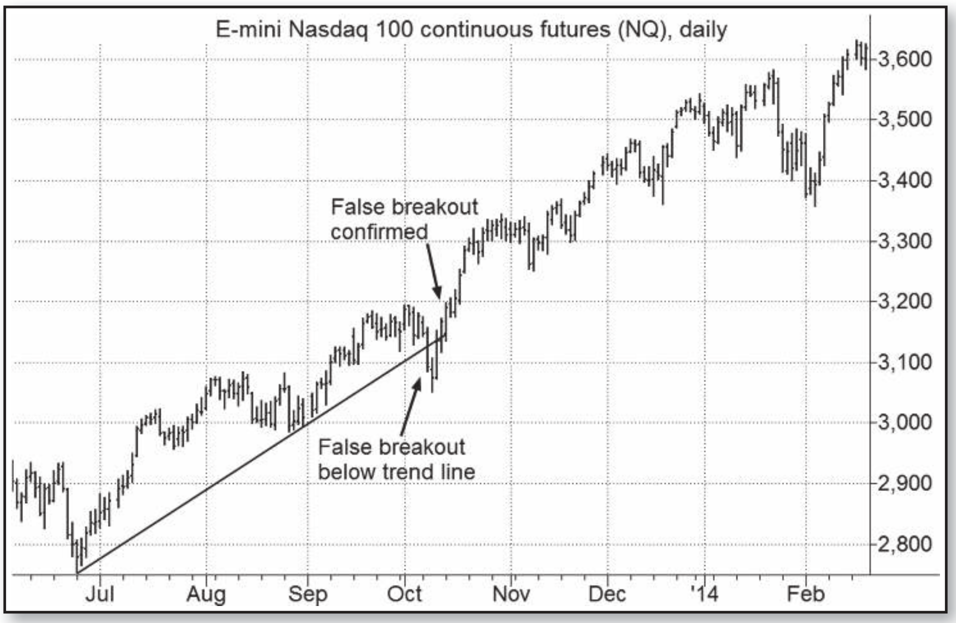

FIGURE 15.10 False Breakout of Uptrend Line: E-Mini Nasdaq Continuous Futures

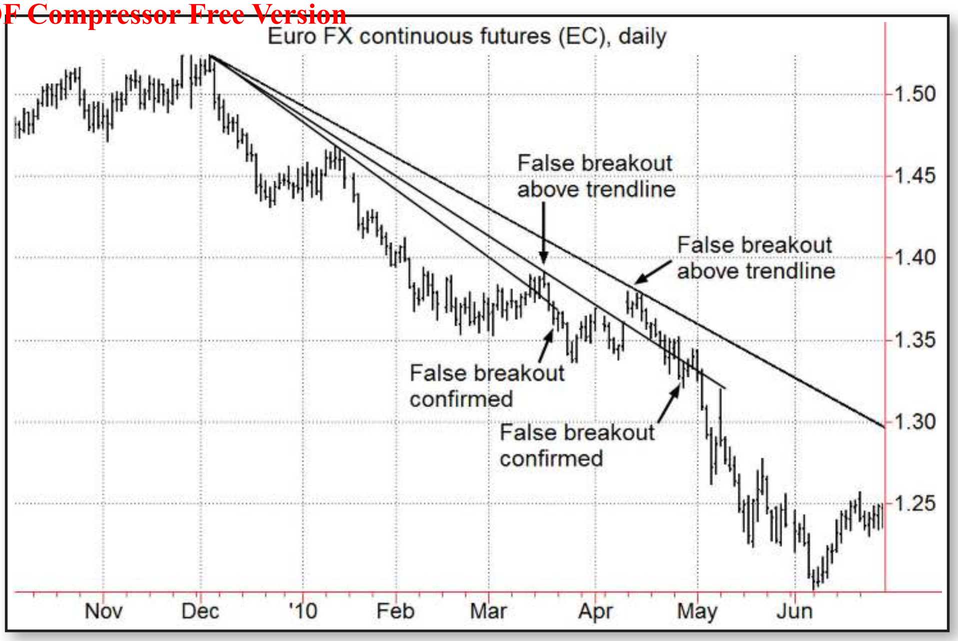

FIGURE 15.11 Multiple False Breakouts of Downtrend Lines: Euro Continuous Futures

Return to Spike Extremes

As was detailed in Chapter 9, price spikes frequently occur at important price reversals. Consequently, the return of prices to a prior spike extreme can be viewed as transforming the original spike into a failed signal. The more extreme the spike (i.e., the greater the magnitude by which the spike high or low exceeds the highs or lows on the prior and subsequent days), the more significant its penetration. The significance of such failed signals is also enhanced if at least several weeks, and preferably several months, have elapsed since the original spike.

In Figure 15.12, the January 2016 return to both the August and October 2015 spike highs was followed by a sharp rally well above the prior spike highs. In Figure 15.13, the October 2010 penetration of the early 2008 spike high was followed by a sharp rally. Figures 15.14 and 15.15 provide two illustrations of downside penetrations of spike lows being followed by steep sell-offs.

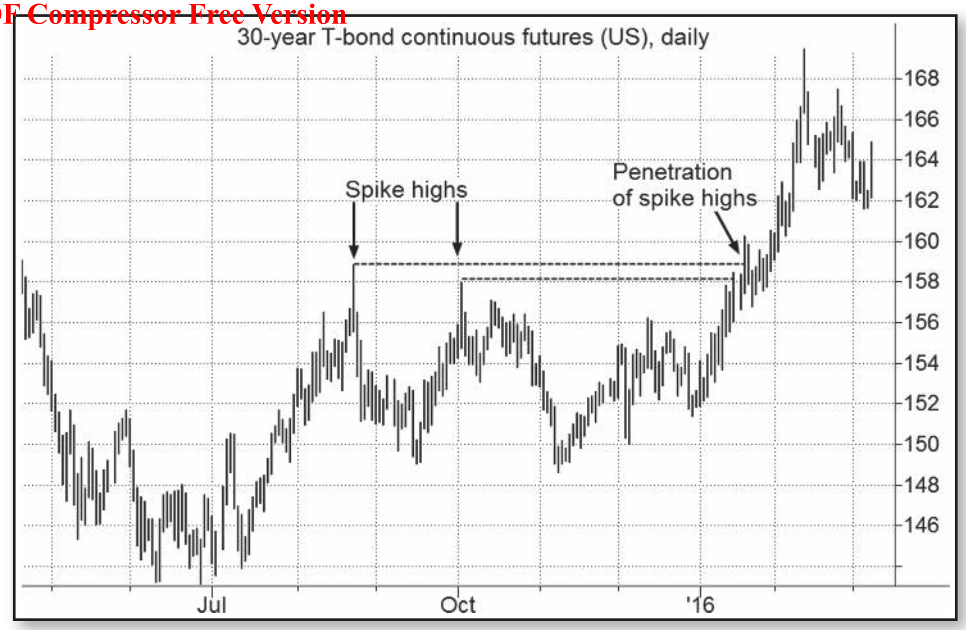

FIGURE 15.12 Penetration of Spike Highs: 30-Year T-Bond Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

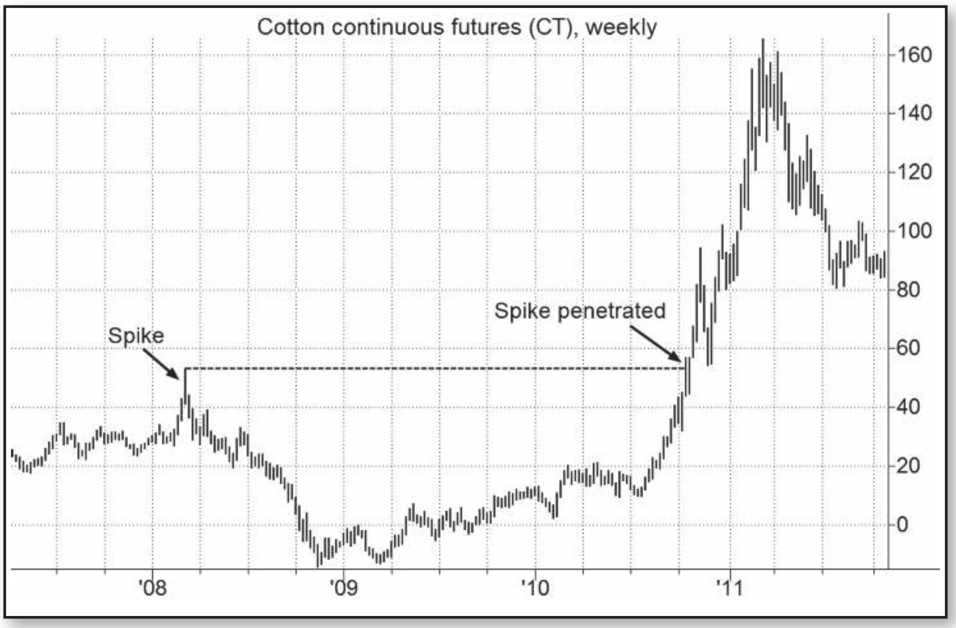

FIGURE 15.13 Penetration of Spike High: Cotton Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

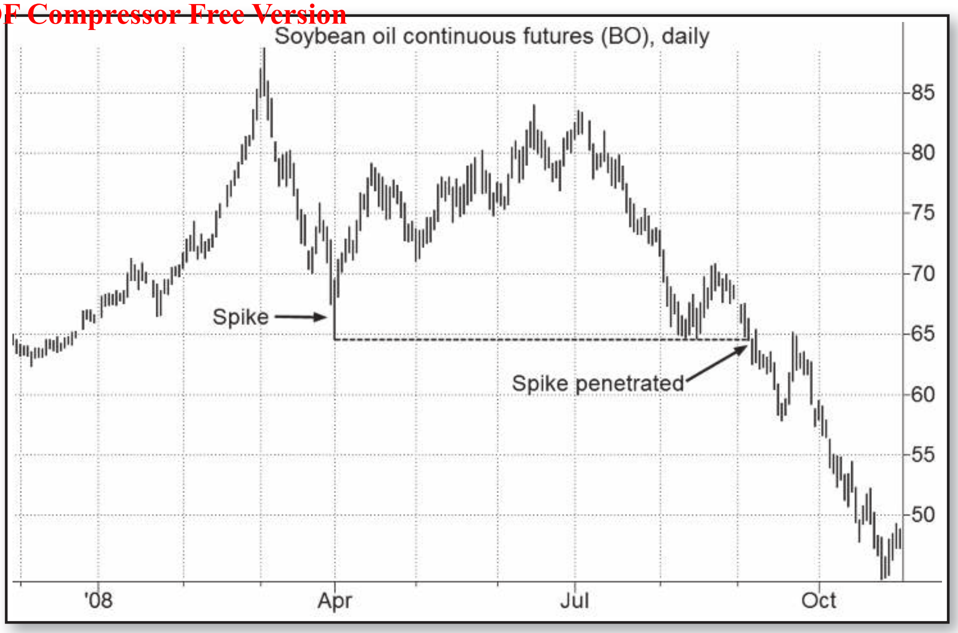

FIGURE 15.14 Penetration of Spike Highs: Soybean Oil Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

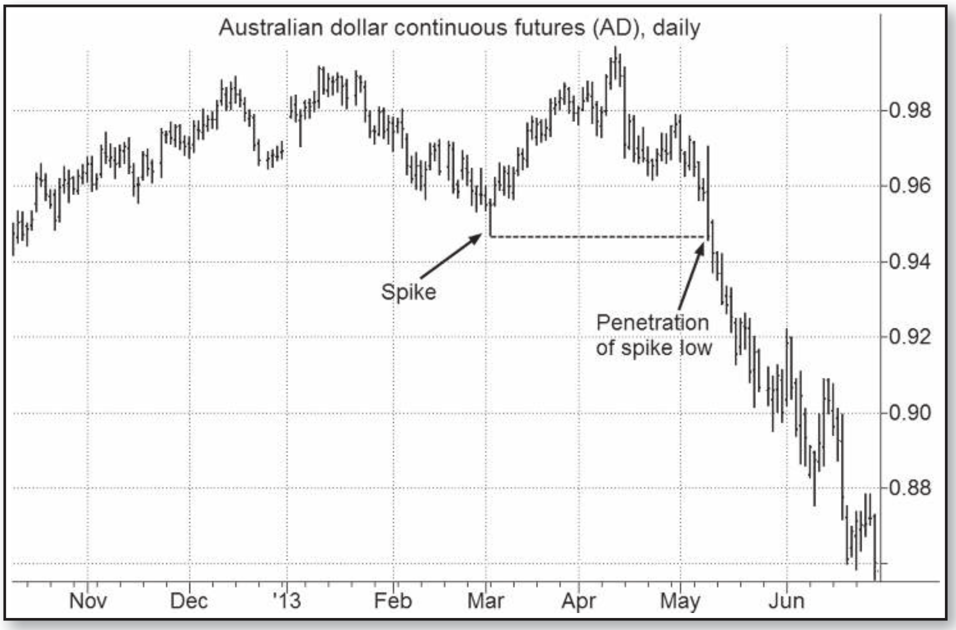

FIGURE 15.15 Penetration of Spike Low: Australian Dollar Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

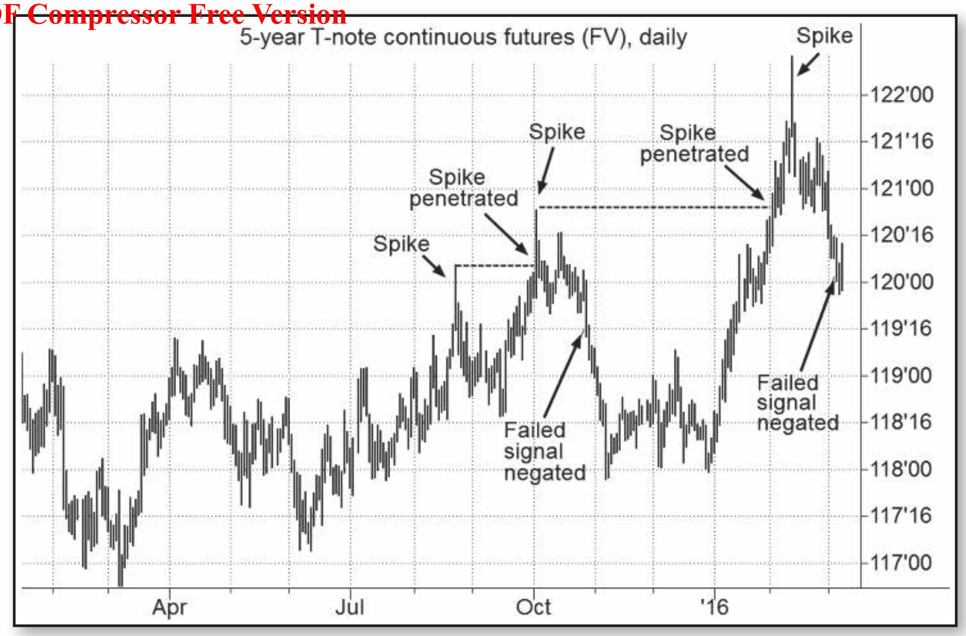

FIGURE 15.16 Spike Penetration Signals Negated: 5-Year T-Note Continuous Futures

Generally speaking, a close beyond the opposite extreme of the spike can be viewed as negating the failed signal. For example, in Figure 15.16 the price briefly exceeded the August spike high, forming a second spike high in early October, but then immediately retreated, falling below the low of the August spike day—a *failed* failed signal, so to speak. This pattern repeated itself in early 2016 when the market penetrated the October spike, rallied for about a week (forming a spike high in the process), but then reversed to close below the low of the October spike day in early March.

Return to Wide-Ranging Day Extremes

As explained in Chapter 9, wide-ranging days (WRDs) with particularly strong or weak closes tend to lead to price extensions in the same direction. Consequently, a close above the high price of a downside WRD or below the low price of an upside WRD can be viewed as confirming such days as failed signals.

In Figure 15.17 the WRD that formed in mid-April 2015 is penetrated to the downside about 10 weeks later, leading to a significant decline. In Figure 15.18 a huge WRD formed in early July 2013 in the vicinity of the May swing high. Three days later, the uptrend was reversed by a downside WRD, which was followed three days later by a close below the low of the first WRD, confirming a failed signal and leading to an extended market slide.

Figure 15.19 shows an example of an up-closing WRD in late April that was reversed by a down-closing WRD 12 days later. A closing penetration of the April WRD occurred four days later and was followed by a large, sustained downtrend. Figure 15.20 shows a massive down-closing WRD that was eclipsed to the upside a little more than a month later and followed by a strong rally to new high ground.

FIGURE 15.17 Penetration of Upside Wide-Ranging Day: Canadian Dollar Continuous Futures

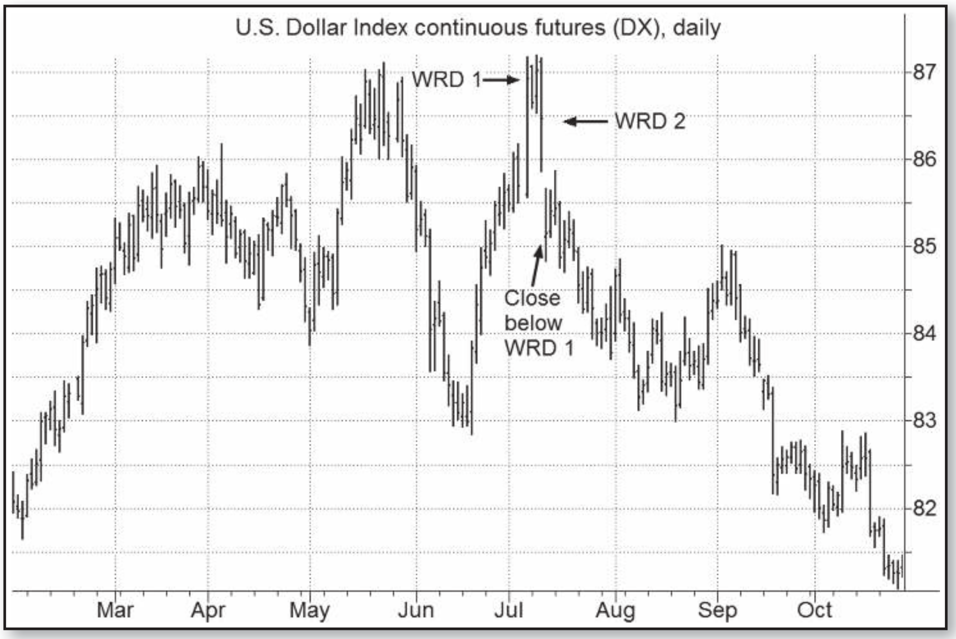

FIGURE 15.18 Penetration of Upside Wide-Ranging Day: U.S. Dollar Index Continuous Futures

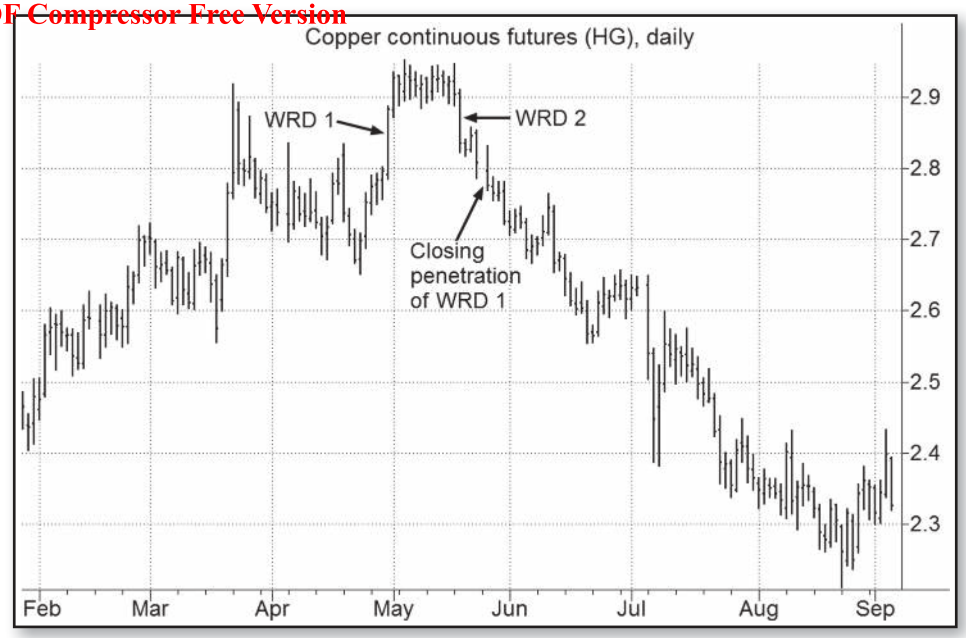

FIGURE 15.19 Penetration of Upside Wide-Ranging Day: Copper Continuous Futures

FIGURE 15.20 Penetration of Downside Wide-Ranging Day: Bund Continuous Futures

Counter-to-Anticipated Breakout of Flag or Pennant

As was explained in Chapter 9, typically, flag or pennant consolidations tend to be followed by price swings in the same direction as the price swings that preceded their formation. Therefore, if a flag or pennant formation is followed by a breakout in the opposite direction of the preceding price swing, it would qualify the pattern as a failed signal.

In Figure 15.21, just as would have been implied by the chart interpretation guidelines presented in Chapter 9, the flag formations that evolved during the 2014 downtrend in soybean prices were generally followed by downswings. The one exception, however, was the flag that formed in late September and early October. In this instance, the flag was followed by an upside breakout. This counter-to-anticipated price action was followed by a rally of more than 13 percent to the mid-November high. Figures 15.22, 15.23, and 15.24 provide three examples where counter-to-anticipated downside breakouts of flag patterns signaled major trend reversals. Note that Figure 15.24 is, in fact, the same reversal depicted in Figure 15.19, which focused on the downside penetration of the strong-closing WRD that immediately preceded the flag. In Figure 15.25 heating oil prices rallied more than 33 percent in one month after the counter-to-anticipated upside breakout of the flag that formed in early 2015.

A counter-to-anticipated breakout does not need to be followed by an immediate extension of the price move in order to be a valid confirmation of a failed signal. How much of a retracement can be allowed before the interpretation of a failed signal is abandoned? One reasonable approach is to

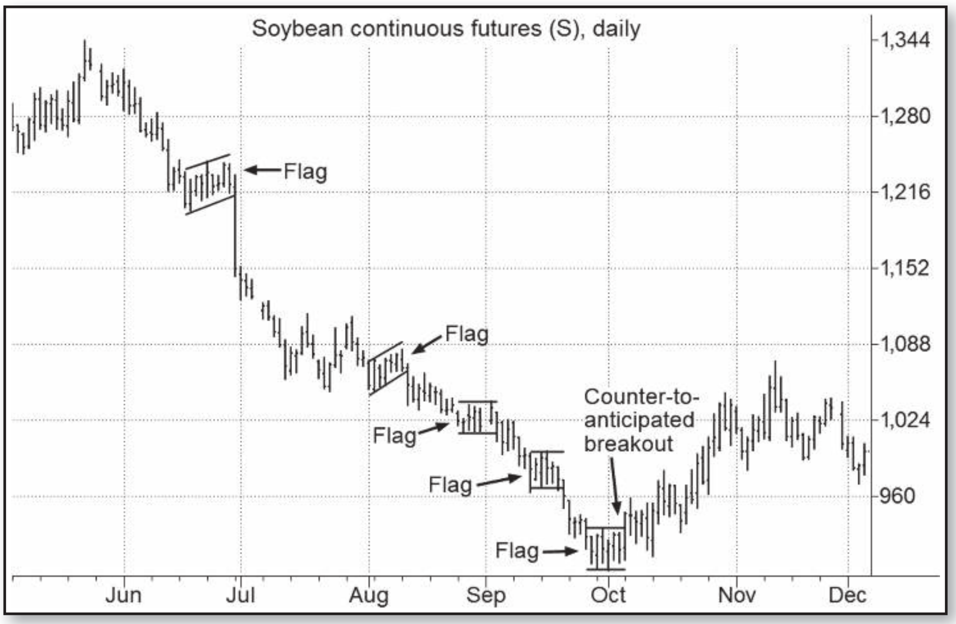

FIGURE 15.21 Counter-to-Anticipated Breakout of Flag Pattern: Soybean Continuous Futures

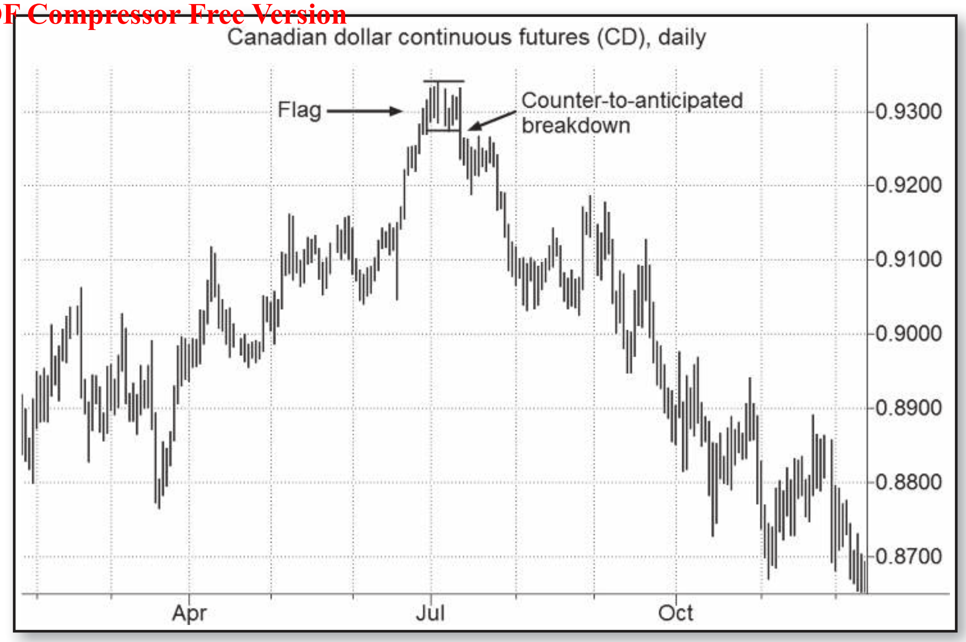

FIGURE 15.22 Counter-to-Anticipated Breakout of Flag Pattern: Canadian Dollar Continuous Futures
Chart created using TradeStation.

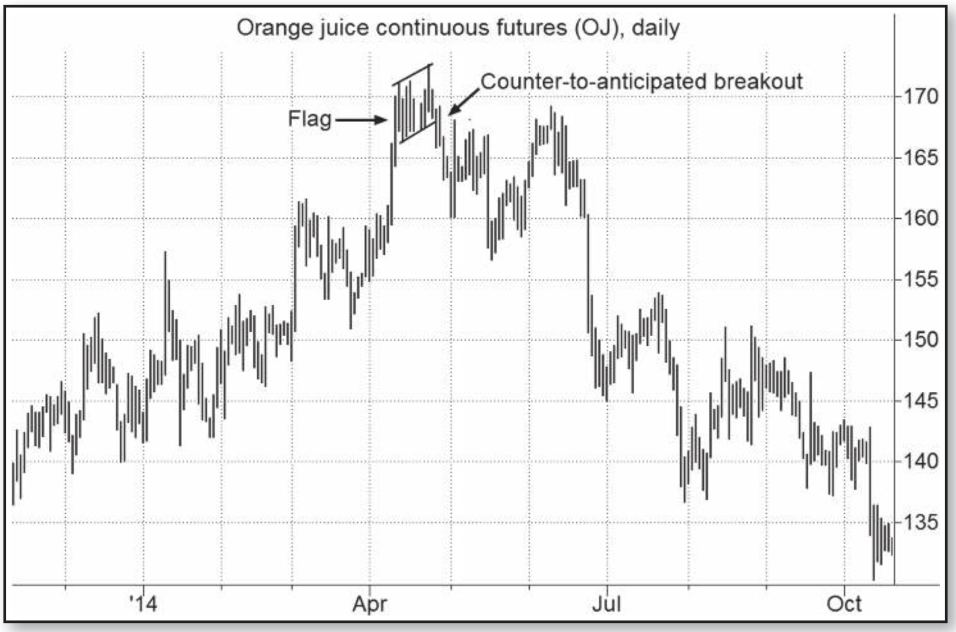

FIGURE 15.23 Counter-to-Anticipated Breakout of Flag Pattern: Orange Juice Continuous Futures
Chart created using TradeStation.

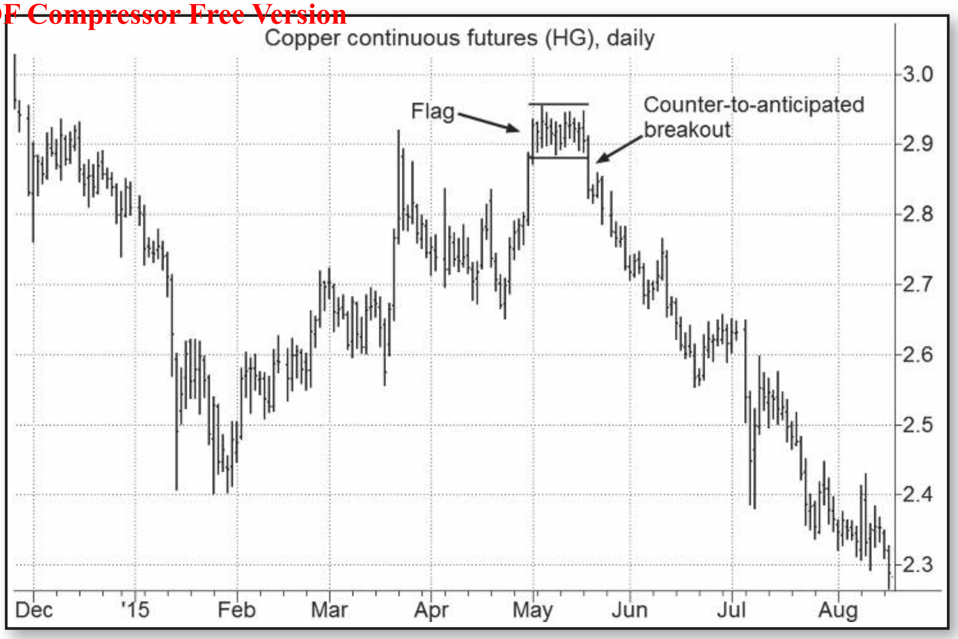

FIGURE 15.24 Counter-to-Anticipated Breakout of Flag Pattern: Copper Continuous Futures

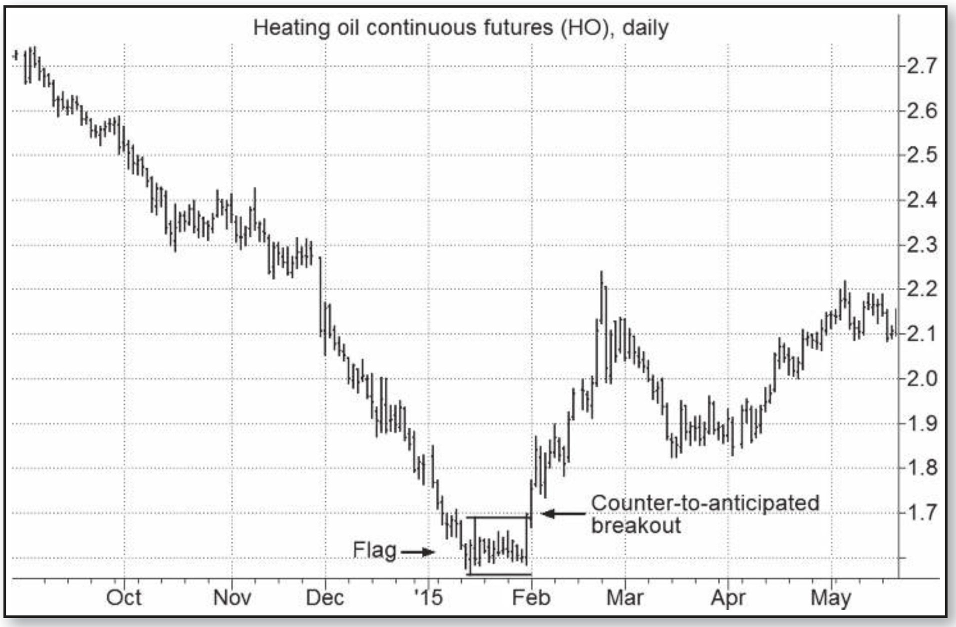

FIGURE 15.25 Counter-to-Anticipated Breakout of Flag Pattern: Heating Oil Continuous Futures

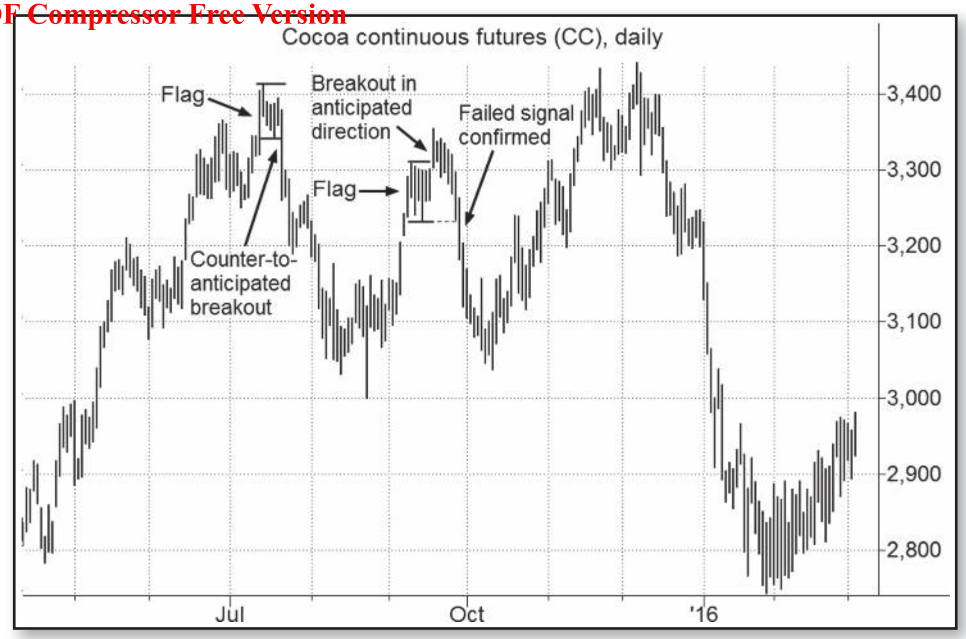

FIGURE 15.26 Counter-to-Anticipated Flag Breakout and Opposite Direction Flag Breakout Following Normal Breakout: Cocoa Continuous Futures

consider the confirmation of a failed signal in force as long as prices do not close beyond the opposite end of the relevant flag or pennant. The retracement in Figure 15.21 provides a good example: after the breakout above the top of the September–October flag, prices pulled back but held at the approximate midpoint of the flag before pushing higher, thereby leaving the failed signal intact.

Figure 15.26 highlights two flag patterns. The first formed in July when prices were rallying and was followed by a sharp sell-off after a counter-to-anticipated breakout to the downside. The second flag occurred in September when the market was rebounding. The market initially broke out of this flag in the expected direction—to the upside—but after a few days prices dropped back into the flag's range and, eventually, penetrated the bottom of the flag, confirming a failed signal pattern. The market subsequently dropped more than 5 percent over the next two weeks. This type of reversal after a normal breakout is the subject of the next section.

Opposite Direction Breakout of Flag or Pennant Following a Normal Breakout

In some cases, flags and pennants are followed by breakouts in the anticipated direction, but prices then reverse to close beyond the opposite extreme of the flag or pennant, as was the case with the September 2015 pattern in Figure 15.26. This combined price action provides another example of

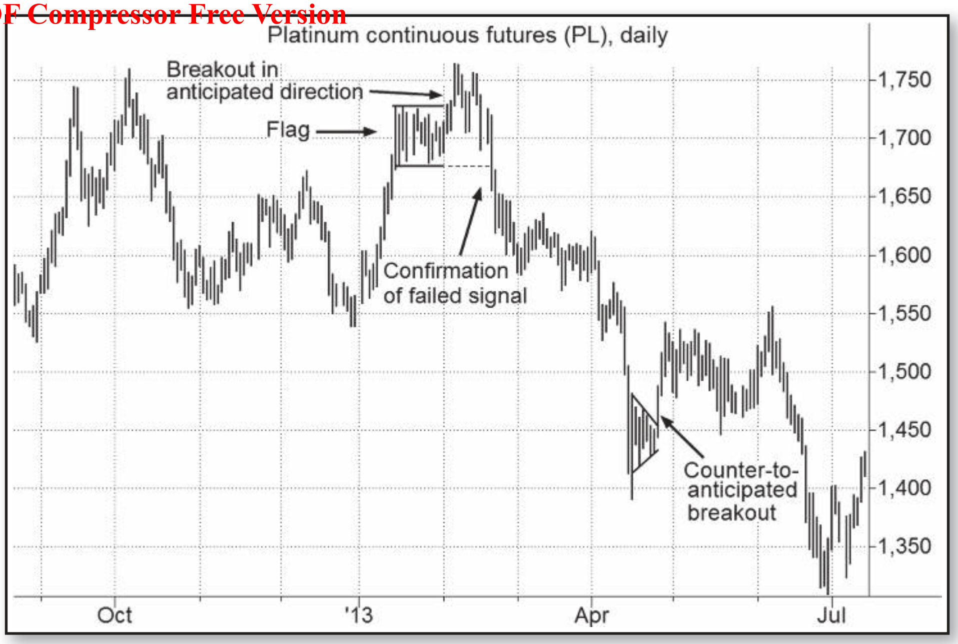

FIGURE 15.27 Opposite Direction Breakout of Flag Following Normal Breakout: Platinum Continuous Futures

a failed signal, since the anticipated breakout of the flag or pennant is followed by a price reversal instead of a price follow-through. Note that a *close* beyond the opposite end of the flag or pennant is required to confirm a failed signal, rather than a mere intraday penetration. Although this more restrictive condition will yield slightly less timely confirmations of failed signals in cases when such a conclusion proves valid, it will reduce the number of inaccurate calls of failed signals.

In Figure 15.27 the flag consolidation that formed in January–February 2013 after an upswing off support near the November–December lows was followed by an upside breakout, as might have been anticipated. Instead of witnessing a further sustained advance, however, prices moved higher for only two days, and less than two weeks later the market had retreated to below the low end of the flag consolidation. This price action qualified the earlier upside breakout above the flag pattern as a failed signal. (Note this type of signal could also be termed a bull or bear trap if it occurs at a major high or low.) In April a counter-to-anticipated upside breakout of a pennant formation was followed by a sharp bounce and consolidation before the market dropped to new lows in June.

In Figure 15.28 the flag that formed during an upswing in natural gas prices was also followed by an upside breakout and then a retreat below the low end of the flag. In this instance, the market pushed back into the flag's range several days later but did not reach the pattern's upper boundary, leaving the failed signal confirmation intact.

Figure 15.29 illustrates a flag pattern that formed during an extended downtrend in sugar futures. The market first broke out of the flag in the anticipated direction but reversed in a few days after

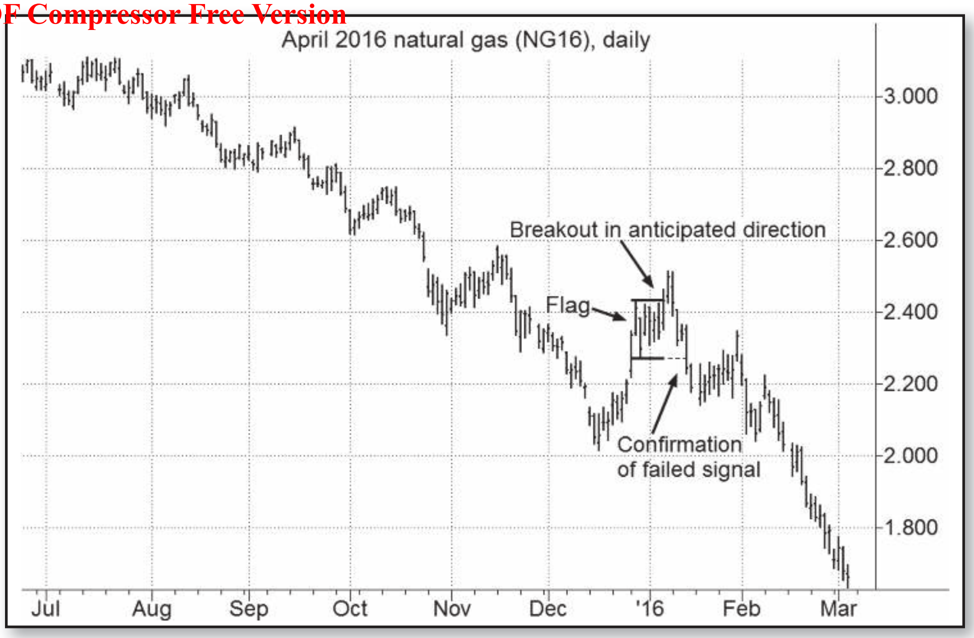

FIGURE 15.28 Opposite Direction Breakout of Flag Following Normal Breakout: April 2016 Natural Gas
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

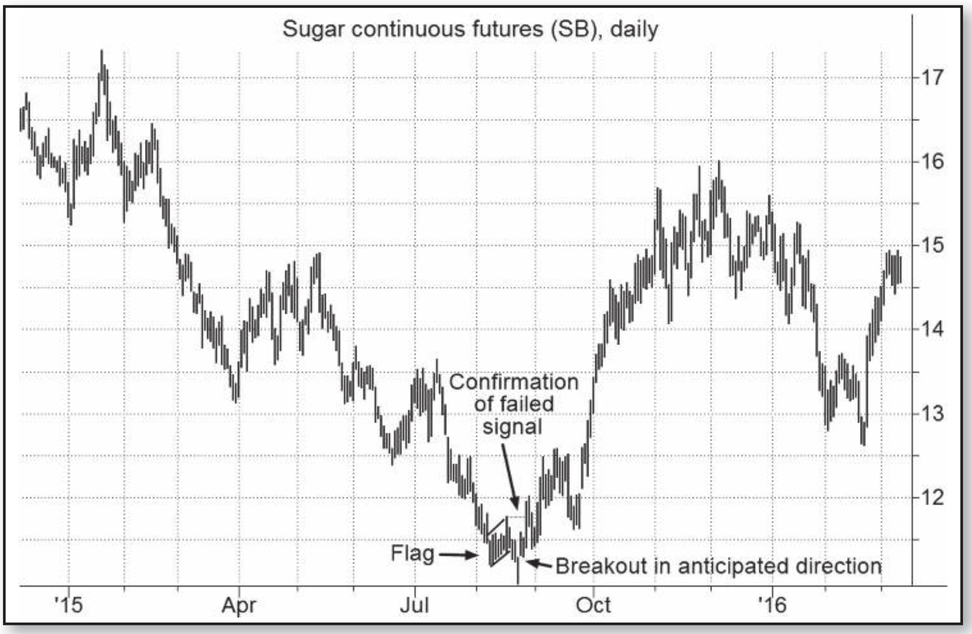

FIGURE 15.29 Opposite Direction Breakout of Flag Following Normal Breakout: Sugar Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

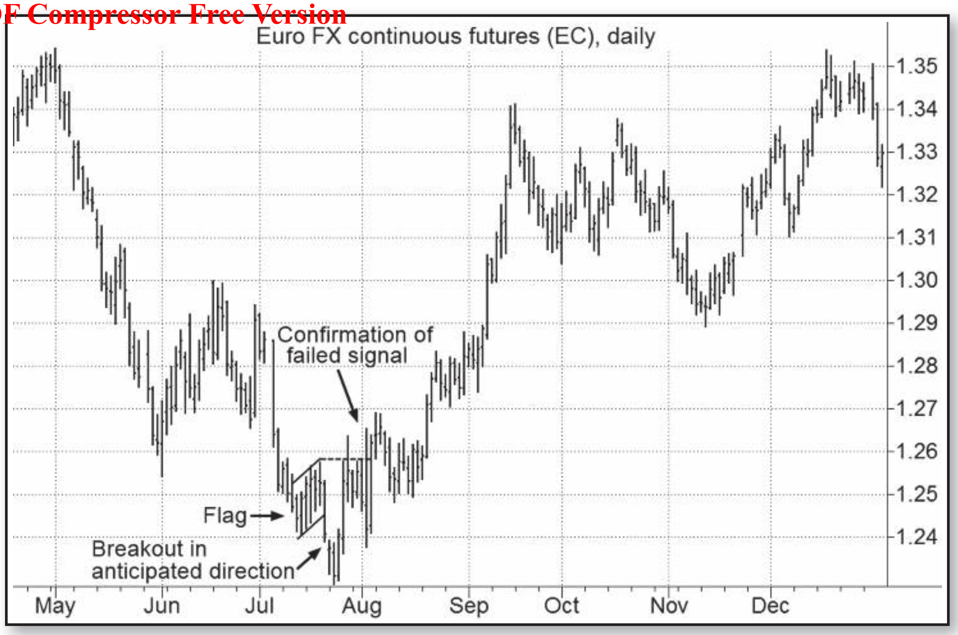

FIGURE 15.30 Opposite Direction Breakout of Flag Following Normal Breakout: Euro Continuous Futures

forming a spike low. The subsequent upside penetration of the flag confirmed the failed signal. After a partial pullback toward the middle of this upward-sloping flag, prices staged a huge upmove. Note this failed signal is also a perfect example of a bear trap bottom.

In Figure 15.30 an expected downside breakout of the flag was followed by an upswing above its upper boundary, confirming a failed flag signal that was followed by a brisk rally.

Penetration of Top and Bottom Formations

The penetration of patterns that are normally associated with major tops and bottoms represents another important type of failed signal. For example, Figure 15.31 illustrates the double top that formed in U.S. 30-year T-bond futures in late 2010 and the penetration of this top several months later. The monthly chart inset shows the extent of the market's subsequent rally. Penetrations of double tops and double bottoms can be significant failure signals even if the top or bottom formation is not confirmed. For example, Figure 15.32 shows the downside penetration of an unconfirmed double bottom—that is, prices did not exceed the pattern's October 2013 intermediate high. Nonetheless, penetration of the pattern's July 2013 and January 2014 lows represented the violation of an important support level, as evidenced by the continued sell-off that followed.

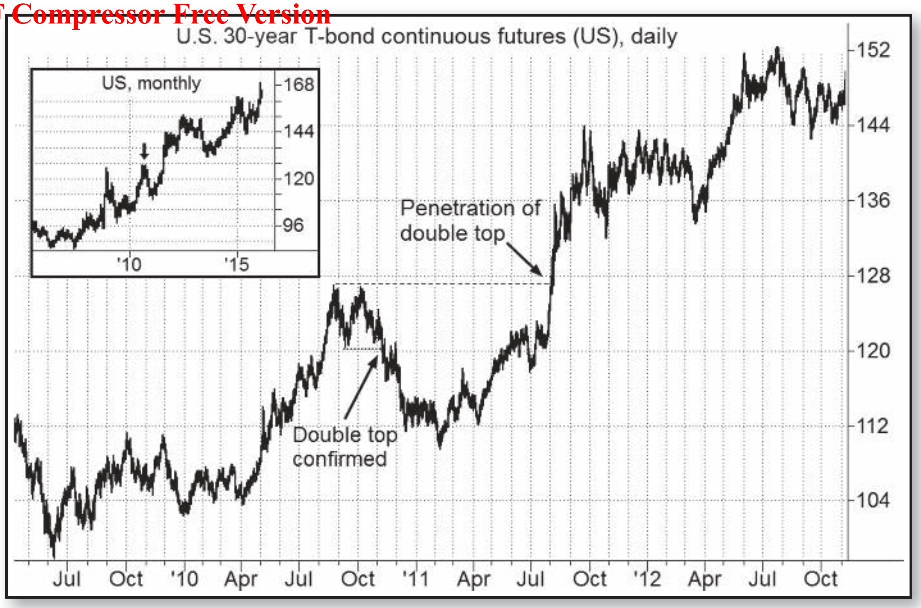

FIGURE 15.31 Penetration of Double Top: 30-Year U.S. T-Bond Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

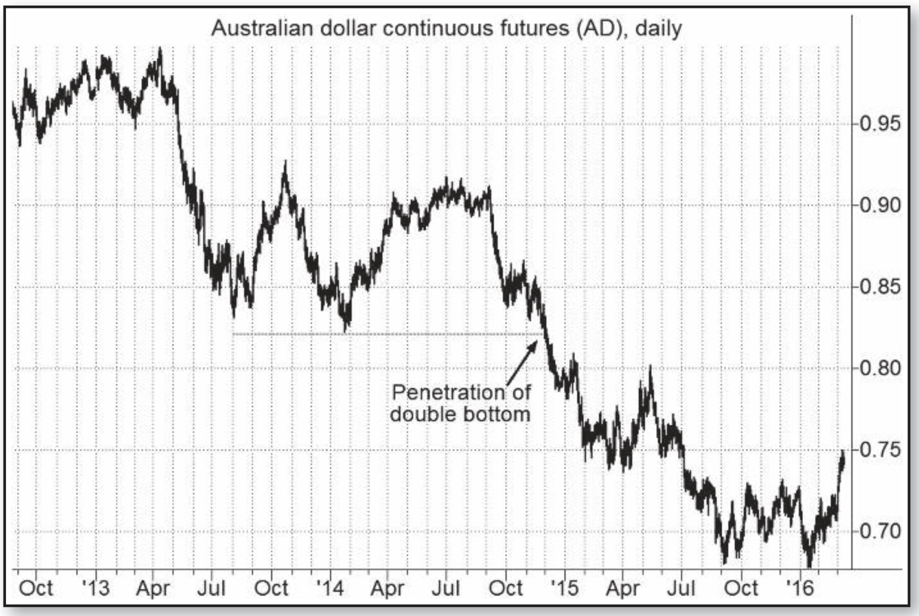

FIGURE 15.32 Penetration of Double Bottom: Australian Dollar Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

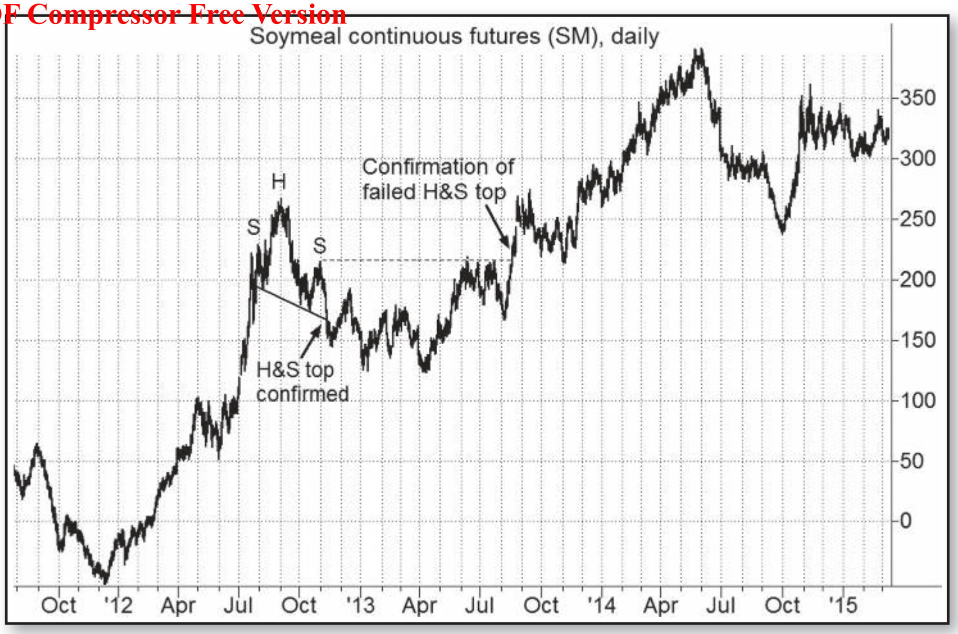

FIGURE 15.33 Failed Head-and-Shoulders Top Pattern: Soymeal Continuous Futures

Penetrations of double-top and double-bottom patterns provide good signals but are relatively rare. Failed signals involving head-and-shoulders patterns are more common and often provide excellent trading indicators. Although the choice of what condition constitutes a confirmation of a failed head-and-shoulders pattern is somewhat arbitrary, I would use the criterion of prices exceeding the most recent shoulder. For example, in Figure 15.33 the rebound above the shoulder that peaked at the beginning of November 2012 would represent a confirmation of a failed head-and-shoulders top pattern. Sometimes prices will first dip back after penetrating the shoulder, even when a substantial advance ultimately ensues, as is the case in Figure 15.34, which shows a long-term example on a weekly chart of the E-mini S&P 500 futures. As long as prices don't close below the relative low formed between the head and right shoulder, the failed signal would remain intact. Figure 15.35 provides another example of a strong rally following a failed head-and-shoulders top.

Figure 15.36 illustrates a failed head-and-shoulders bottom pattern. In analogous fashion to the head-and-shoulders top case, the downside penetration of the more recent shoulder is used as the confirmation condition of a failed signal.

The trader may often benefit by waiting for a retracement before implementing a position based on the confirmation of a failed head-and-shoulders pattern, as illustrated by Figure 15.34. The tradeoff is that such a strategy will result in missing very profitable trades in those cases where there is no retracement or only a very modest retracement (e.g., Figures 15.35 and 15.36).

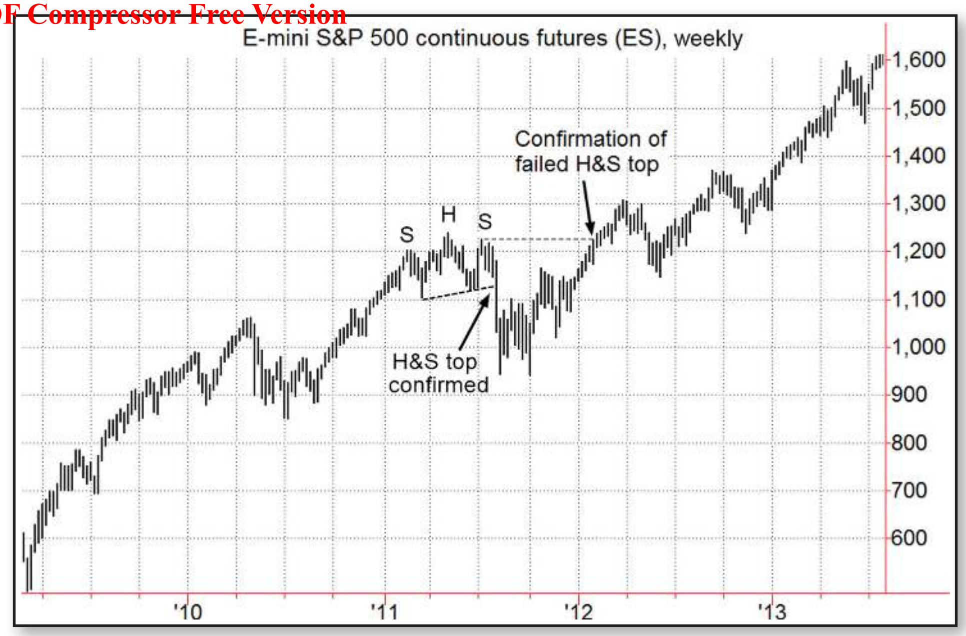

FIGURE 15.34 Failed Head-and-Shoulders Top Pattern: E-Mini S&P 500 Continuous Futures

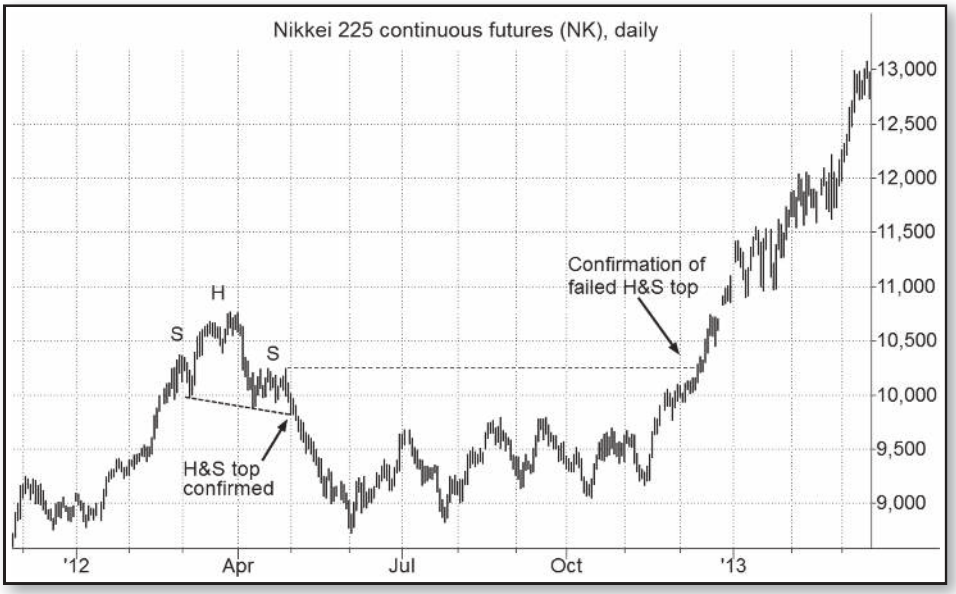

FIGURE 15.35 Failed Head-and-Shoulders Top Pattern: Nikkei 225 Continuous Futures

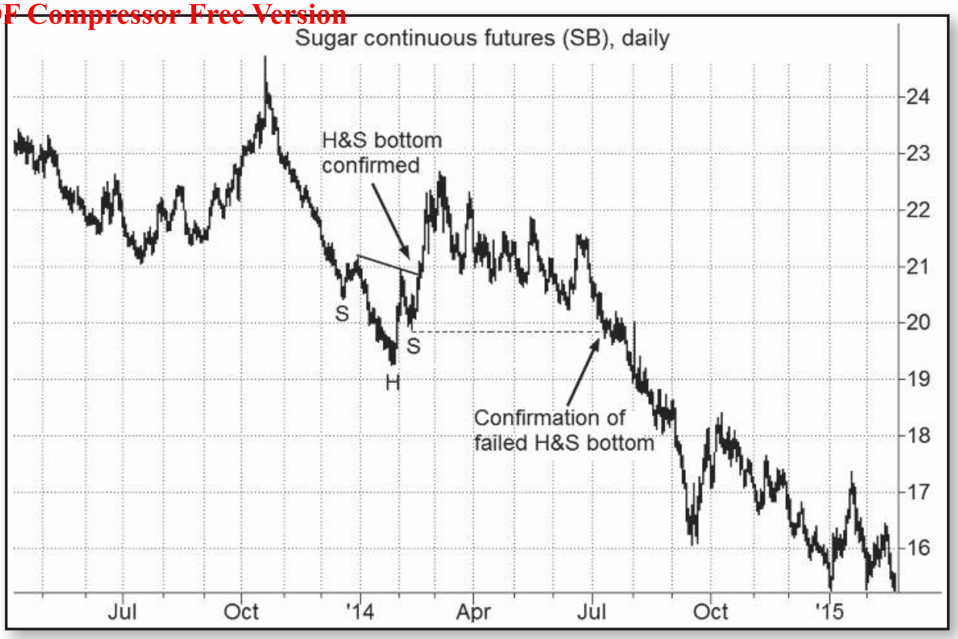

FIGURE 15.36 Failed Head-and-Shoulders Bottom Pattern: Sugar Continuous Futures

Breaking of Curvature

As was discussed in Chapter 9, rounding patterns often provide very reliable trading signals. In this sense, the breaking of a curved price pattern can be viewed as transforming the pattern into a failed signal. Figure 15.37 actually contains two examples where the breaking of the curvature of what had been an apparent rounding-top pattern represented a bullish signal. In Figure 15.38, the breaking of the curvature of an apparent rounding-bottom pattern led to a steep decline in corn prices in 2014. Note the downthrust in January 2014 (the low of the curved pattern) was the bear trap illustrated in Figures 15.4 and 15.8. So, in effect, this chart illustrates two successive failure patterns, the first signaling a near-two-month rebound, and the second the subsequent reversal into a major downtrend.

The Future Reliability of Failed Signals

There is an inverse relationship between the popularity of an indicator and its efficiency. For example, decades ago, when technical analysis was used by fewer market practitioners, chart breakouts (price moves above or below prior trading ranges) tended to work relatively well, providing many excellent signals without an abundance of false signals. In my observation, as technical analysis became increasingly popular and breakouts a commonly used tool, the efficiency of this pattern seemed to deteriorate. In fact, it now seems that price *reversals* following breakouts may more often be the rule than the exception.

FIGURE 15.37 Breaking of Curvature: E-Mini Nasdaq 100 Continuous Futures

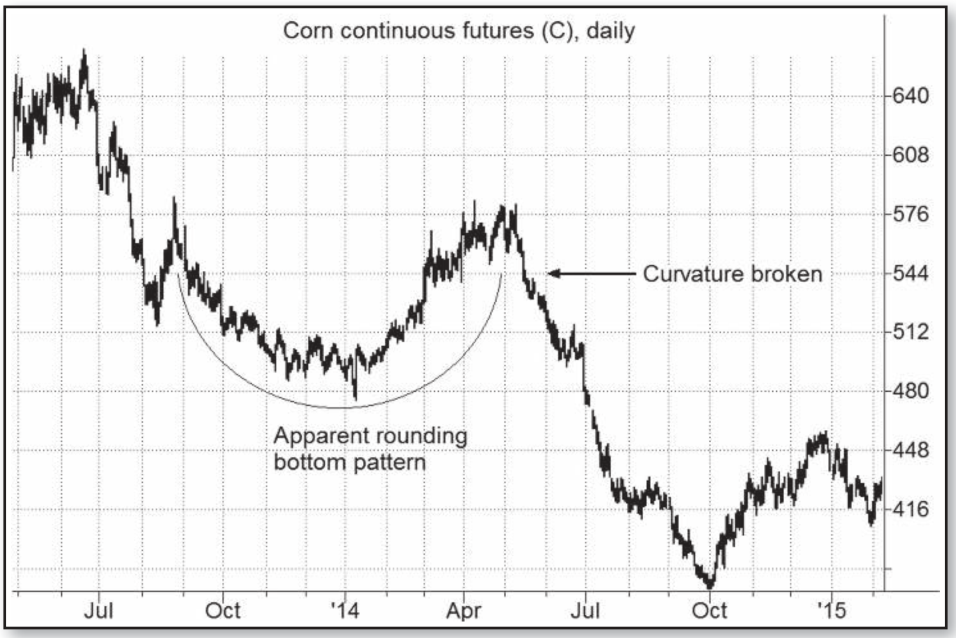

FIGURE 15.38 Breaking of Curvature: Corn Continuous Futures

As stated earlier, I find failed signals considerably more reliable than conventional chart patterns. Although the concept of failed signals is certainly not new, I don't believe its usage is widely emphasized. If the use of failed signals were to become significantly more widespread, however, their long-term reliability could be adversely affected.

As a final comment, it should be emphasized that the concept of failed signals in this chapter has been presented in the context of conventional chart analysis as it exists today. In the future—particularly the distant future—what passes for popular chart interpretation may well change. The concept of failed signals, however, can be made dynamic by pegging it to the conventional wisdom. In other words, if a new chart pattern became popular as a technical signal in the future (e.g., in the way breakouts are widely used today), a failure of the pattern could be viewed as more significant than the pattern itself. In this more general sense, the concept of failed signals could prove timeless.

Conclusion

The novice trader will ignore a failed signal, riding a position into a large loss while hoping for the best. More experienced traders, having learned the importance of money management, will exit quickly once it is apparent they have made a bad trade. However, the truly skilled trader will be able to do a 180-degree turn, reversing a position at a loss if market behavior (e.g., confirmation of a failed signal) points to such a course of action. In other words, it takes great discipline to capitalize on failed signals, but such flexibility is essential to the effective synthesis of chart analysis and trading.

PART IV

Trading Systems and Performance Measurement

CHAPTER 16

Technical Trading Systems: Structure and Design

There are only two types of trend-following systems: fast and slow.

—Jim Orcutt

Be forewarned. If you are expecting to find the blueprint for a heretofore secret trading system that *consistently* makes 100 percent plus per year *in real-life trading* with minimal risk, you'll have to look elsewhere. For one thing, I have not yet discovered such a "sure thing" money machine. But, in a sense, that is beside the point. Quite frankly, I have always been somewhat puzzled by advertisements for books or computer software promising to reveal the secrets of systems that make 100 percent, 200 percent, and more! Why are they selling such valuable information for $99, or even $2,999?

The primary goal of this chapter is to provide readers with the background knowledge necessary to develop their own trading systems. The discussion focuses on the following five areas:

1. An overview of some basic trend-following systems
2. The key weaknesses of these systems
3. Guidelines for transforming "generic" systems into more powerful systems
4. Countertrend systems
5. Diversification as a means of improving performance

Chapter 17 provides additional examples of trading systems, using original systems as illustrations. The essential issues of appropriate data selection, system testing procedures, and performance measurement are discussed in Chapters 18, 19, and 20.

The Benefits of a Mechanical Trading System

Is paper trading easier than real trading? Most speculators would answer yes, even though both tasks require an equivalent decision process. This difference is explained by a single factor: emotion. Overtrading, premature liquidation of good positions because of rumors, jumping the gun on market entry to get a better price, riding a losing position—these are but a few of the negative manifestations of emotion in actual trading. Perhaps the greatest value of a mechanical system is that it eliminates emotion from trading. In so doing, it allows the speculator to avoid many of the common errors that often impede trading performance. Furthermore, removing the implied need for constant decision making substantially reduces trading-related stress and anxiety.

Another benefit of a mechanical system is that it ensures a consistent approach—that is, the trader follows all signals indicated by a common set of conditions. This is important, since even profitable trading strategies can lose money if applied selectively. To illustrate this point, consider the example of a market advisory whose recommendations yield a net profit over the long run (after allowances for commissions and poor executions). Will the advisory's subscribers make money if they only implement trades in line with its recommendations? Not necessarily. Some people will pick and choose trades, invariably missing some of the biggest winners. Others will stop following the recommendations after the advisor has a losing streak, and as a result may miss a string of profitable trades. The point is that a good trading strategy is not sufficient; success also depends on consistency.

A third advantage of mechanical trading systems is they normally provide the trader with a method for controlling risk. Money management is an essential ingredient of trading success. Without a plan for limiting losses, a single bad trade can lead to disaster. Any properly constructed mechanical system will either contain explicit stop-loss rules or specify conditions for reversing a position given a sufficient adverse price move. As a result, following signals generated by a mechanical trading system will normally prevent the possibility of huge losses on individual trades (except in extreme circumstances when one is unable to liquidate a position because the market is in the midst of a string of locked-limit moves). Thus, the speculator using a mechanical system may end up losing money due to the cumulative effect of a number of negative trades, but at least his account will not be decimated by one or two bad trades.

Of course, money management does not necessarily require the use of a trading system. Risk control can also be achieved by initiating a good-till-canceled stop order whenever a new position is taken, or by predetermining the exit point upon entering a trade and sticking to that decision. However, many traders lack sufficient discipline and will be tempted to give the market just a little more time once too often.

Three Basic Types of Systems

The categories used to classify trading systems are completely arbitrary. The following three-division classification is intended to emphasize a subjective interpretation of the key conceptual differences in possible trading approaches:

Trend-following. A trend-following system waits for a specified price move and then initiates a position in the same direction based on the implicit assumption that the trend will continue.

Countertrend. A countertrend system waits for a significant price move and then initiates a position in the opposite direction on the assumption that the market is due for a correction.

Pattern recognition. In a sense, all systems can be classified as pattern recognition systems. After all, the conditions that signal a trend or a countertrend trade are a type of pattern (e.g., close beyond the 20-day high or low). However, the implication here is that the chosen patterns are not based primarily on directional moves, as is the case in trend-following and countertrend systems. For example, a pattern-recognition system might generate signals on the basis of "spike days" (see Chapter 9). In this case, the key consideration is the pattern itself (e.g., spike) rather than the extent of any preceding price move. Of course, this example is overly simplistic. In practice, the patterns used for determining trading signals will be more complex, and several patterns may be incorporated into a single system.

Systems of this type may sometimes employ probability models in making trading decisions. In this case the researcher would try to identify patterns that appeared to act as precursors of price advances or declines in the past. An underlying assumption in this approach is that such past behavioral patterns can be used to estimate current probabilities for rising or declining markets given certain specified conditions. This chapter does not elaborate on this approach of trading system design since it lies beyond the scope of the overall discussion.

It should be emphasized that the lines dividing the preceding categories are not always clear-cut. As modifications are incorporated, a system of one type may begin to more closely approximate the behavioral pattern of a different system category.

Trend-Following Systems

By definition, trend-following systems never sell near the high or buy near the low, because a meaningful opposite price move is required to signal a trade. Thus, in using this type of system, the trader will always miss the first part of a price move and may surrender a significant portion of profits before an opposite signal is received (assuming the system is always in the market). There is a basic trade-off involved in the choice of the sensitivity, or speed, of a trend-following system. A sensitive system, which responds quickly to signs of a trend reversal, will tend to maximize profits on valid signals, but it will also generate far more false signals. A nonsensitive, or slow, system will reflect the reverse set of characteristics.

Many traders become obsessed with trying to catch every market wiggle. Such a predilection leads them toward faster and faster trend-following systems. Although in some markets fast systems consistently outperform slow systems, in most markets the reverse is true, as the minimization of losing trades and commission costs in slow systems more than offsets the reduced profits in the good trades. This observation is only intended as a cautionary note against the natural tendency toward seeking out more sensitive systems. However, in all cases, the choice between fast and slow systems must be determined on the basis of empirical observation and the trader's subjective preferences.

There is a wide variety of possible approaches in constructing a trend-following system. In this chapter we focus on two of the most basic methods: moving average systems and breakout systems.

Moving Average Systems

The moving average for a given day is equal to the average of that day's closing price and the closing prices on the preceding $N - 1$ days, where N is equal to the number of days in the moving average.

For example, in a 10-day moving average, the appropriate value for a given day would be the average of the 10 closing prices culminating with that day. The term *moving average* refers to the fact that the set of numbers being averaged is continuously moving through time.

Because the moving average is based on past prices, in a rising market the moving average will be below the price, while in a declining market the moving average will be above the price. Thus, when a price trend reverses from up to down, prices must cross the moving average from above. Similarly, when the trend reverses from down to up, prices must cross the moving average from below. In the most basic type of moving average system, these crossover points are viewed as trade signals: a buy signal is indicated when prices cross the moving average from below; a sell signal is indicated when prices cross the moving average from above. The crossover should be determined based on closing prices. Table 16.1 illustrates the calculation of a 10-day simple moving average and indicates the corresponding crossover signal points.

TABLE 16.1 Calculating a Moving Average

Day	Closing Price	10-Day Moving Average	Crossover Signal
1	80.50		
2	81.00		
3	81.90		
4	81.40		
5	83.10		
6	82.60		
7	82.20		
8	83.10		
9	84.40		
10	85.20	82.54	
11	84.60	82.95	
12	83.90	83.24	
13	84.40	83.49	
14	85.20	83.87	
15	86.10	84.17	
16	85.40	84.45	
17	84.10	84.64	Sell
18	83.50	84.68	
19	83.90	84.63	
20	83.10	84.42	
21	82.50	84.21	
22	81.90	84.01	
23	81.20	83.69	
24	81.60	83.33	
25	82.20	82.94	
26	82.80	82.68	Buy
27	83.40	82.61	
28	83.80	82.64	
29	83.90	82.64	
30	83.50	82.68	

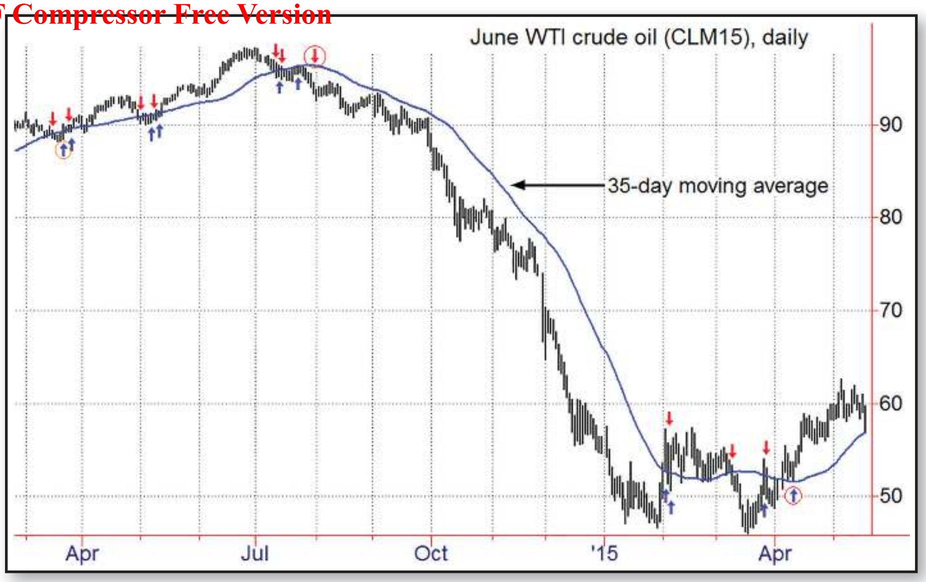

FIGURE 16.1 June 2015 WTI Crude Oil and 35-Day Moving Average
Note: ↑ = buy signal: prices cross moving average from below and close above line; ↓ = sell signal: prices cross moving average from above and close below line; Ⓘ = buy signal not eliminated by filter; Ⓘ = sell signal not eliminated by filter.

Figure 16.1 shows the June 2015 WTI crude oil contract with a 35-day moving average. The non-circled buy and sell signals on the chart are based on the simple moving average system just described. (For now ignore the circled signals; they are explained later.) Note that although the system catches the major downtrend, it also generates several false signals. Of course, this problem can be mitigated by increasing the length of the moving average, but the tendency toward excessive false signals is a characteristic of the simple moving average system. The reason for this is that temporary, sharp price fluctuations, sufficient to trigger trade signals, are commonplace events in futures markets.

One school of thought suggests the problem with the simple moving average system is that it weights all days equally, whereas more recent days are more important and hence should be weighted more heavily. Many different weighting schemes have been proposed for constructing moving averages. Two of the most common weighting approaches are the *linearly weighted moving average* (LWMA) and the *exponentially weighted moving average* (EWMA).[1]

The LWMA assigns the oldest price in the moving average a weight of 1, the second oldest price a weight of 2, and so on. The weight of the most recent price would be equal to the number of days

[1]The following two sources were used as reference for the remainder of this section: (1) Perry Kaufman, *Trading Systems and Methods* (Hoboken, NJ: John Wiley & Sons, 2013), and (2) *Technical Analysis of Stocks & Commodities,* bonus issue 1995, sidebar, page 66.

in the moving average. The LWMA is equal to the sum of the weighted prices divided by the sum of the weights:

$$\text{LWMA} = \frac{\sum_{t=1}^{n} P_t \cdot t}{\sum_{t=1}^{n} t}$$

where t = time indicator (oldest day = 1, second oldest = 2, etc.)
P_t = price at time t
n = number of days in moving average

For example, for a 10-day LWMA, the price of 10 days ago would be multiplied by 1, the price of 9 days ago by 2, and so on through the most recent price, which would be multiplied by 10. The sum of these weighted prices would then be divided by 55 (the sum of 1 through 10) to obtain the LWMA.

The EWMA is calculated as the sum of the current price multiplied by a *smoothing constant* between 0 and 1, denoted by the symbol a, and the previous day's EWMA multiplied by $1 - a$:

$$\text{EWMA}_t = aP_t + (1 - a)\text{EWMA}_{t-1}$$

This linked calculation wherein each day's value of the EWMA is based on the previous day's value means that *all* prior prices will have some weight, but the weight of each day drops exponentially the further back in time it is. The weight of any individual day would be:

$$a(1 - a)^k$$

where k = number of days prior to current day (for current day, $k = 0$ and term reduces to a).

Since a is a value between 0 and 1, the weight of each given day drops sharply moving back in time. For example, if $a = 0.1$, yesterday's price would have a weight of 0.09, the price two days ago would have a weight of 0.081, the price 10 days ago would have a weight of 0.035, and the price 30 days ago would have a weight of 0.004.

An EWMA with a smoothing constant, a, corresponds roughly to a simple moving average of length n, where a and n are related by the following formula:

$$a = 2/(n + 1)$$

or

$$n = (2 - a)/a$$

Thus, for example, an EWMA with a smoothing constant equal to 0.1 would correspond roughly to a 19-day simple moving average. As another example, a 40-day simple moving average would correspond roughly to an EWMA with a smoothing constant equal to 0.04878.

In my view, there is no strong empirical evidence to support the idea that linearly or exponentially weighted moving averages provide a substantive and consistent improvement over simple moving averages. Sometimes weighted moving averages will do better; sometimes simple moving averages will do better. (See Chapter 11 for an illustration of this point.) The question of which method will yield better results will be entirely dependent on the markets and time periods selected, with no reason to assume that *past* relative superiority will be indicative of the probable *future* pattern. In short, experimentation with different weighted moving averages probably does not represent a particularly fruitful path for trying to improve the simple moving average system.

A far more meaningful improvement is provided by the crossover moving average approach. In this system, trade signals are based on the interaction of two moving averages, as opposed to the interaction between a single moving average and price. The trading rules are very similar to those of the simple moving average system: a buy signal is generated when the shorter moving average crosses above the longer moving average; a sell signal is generated when the shorter moving average crosses below the longer moving average. (In a sense, the simple moving average system can be thought of as a special case of the crossover moving average system, in which the short-term moving average is equal to 1.) Because trade signals for the crossover system are based upon two smoothed series (as opposed to one smoothed series and price), the number of false signals is substantially reduced. Figures 16.2, 16.3, and 16.4 compare trade signals generated by a simple 12-day moving average system, a simple 48-day moving average system, and the crossover system based on these two averages. Generally speaking, the crossover moving average system is far superior to the simple moving average. (However, it should be noted that

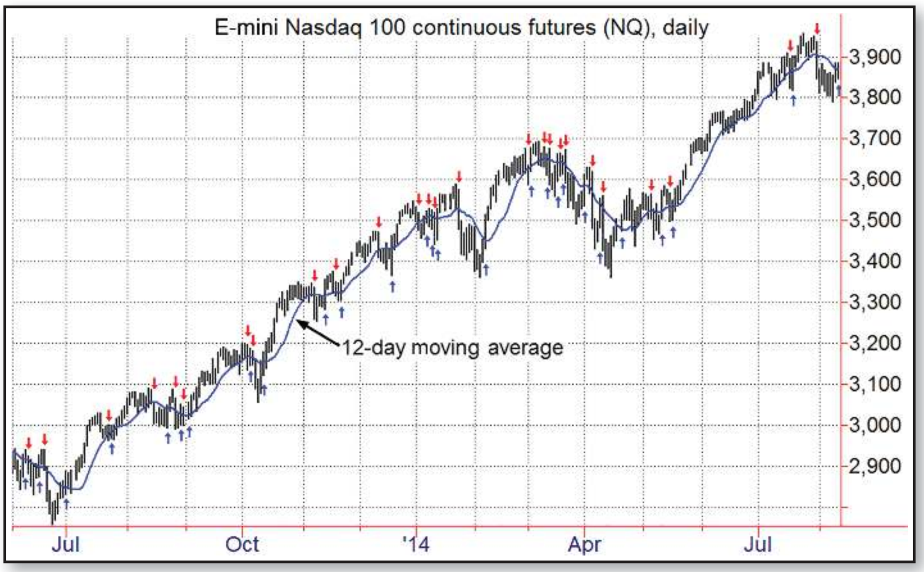

FIGURE 16.2 E-Mini Nasdaq 100 Continuous Futures with 12-Day Moving Average
Note: ↑ = buy signal: prices cross moving average from below and close above line; ↓ = sell signal: prices cross moving average from above and close below line.
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

FIGURE 16.3 E-Mini Nasdaq 100 Continuous Futures with 48-Day Moving Average
Note: ↑ = buy signal: prices cross moving average from below and close above line; ↓ = sell signal: prices cross moving average from above and close below line.

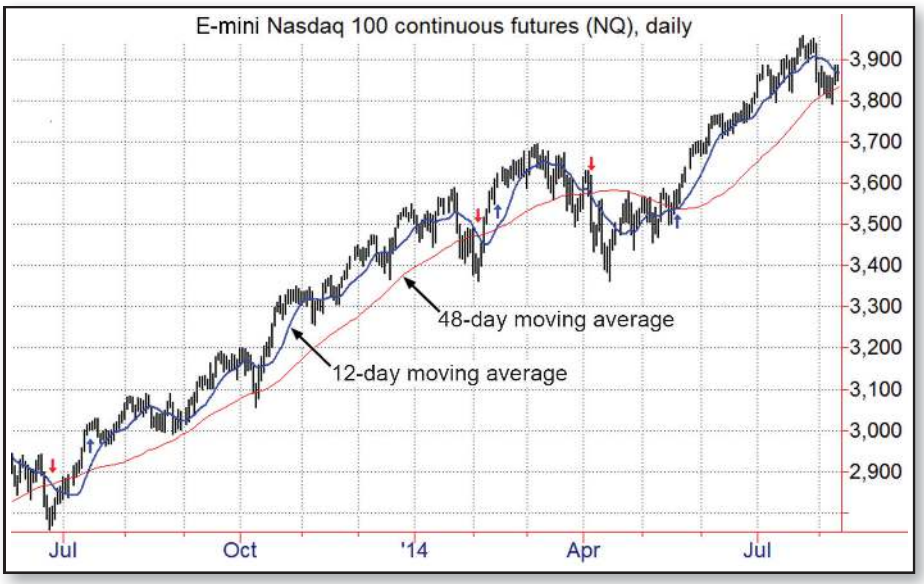

FIGURE 16.4 E-Mini Nasdaq 100 Continuous Futures with Moving Average Crossover
Note: ↑ = buy signal: short-term moving average (12-day) crosses long-term moving average (48-day) from below; ↓ = sell signal: short-term moving average crosses long-term moving average from above.

by including some of the trend-following-system modifications discussed in a later section, even the simple moving average system can provide the core for a viable trading approach.) The weaknesses of the crossover moving average system and possible improvements are discussed later.

Breakout Systems

The basic concept underlying breakout systems is very simple: the ability of a market to move to a new high or low indicates the potential for a continued trend in the direction of the breakout. The following set of rules provides an example of a simple breakout system:

1. Cover short and go long if today's close exceeds the prior *N*-day high.
2. Cover long and go short if today's close is below the prior *N*-day low.

The value chosen for *N* will define the sensitivity of the system. If a short-duration period is used for comparison to the current price (e.g., $N = 7$), the system will indicate trend reversals fairly quickly, but will also generate many false signals. In contrast, the choice of a longer-duration period (e.g., $N = 40$) will reduce false signals, but at the cost of slower entry.

Figure 16.5 compares the trade signals generated by the preceding simple breakout system in silver continuous futures using $N = 7$ and $N = 40$. The following three observations, which are evidenced in Figure 16.5, are also valid as generalizations describing the trade-offs between fast and slow breakout systems:

1. A fast system will provide an earlier signal of a major trend transition (e.g., the October 2012 sell signal).

FIGURE 16.5 Breakout System Signals, Fast versus Slow Systems: Silver Continuous Futures
Note: B, S = signals for $N = 7$; B, S = signals for $N = 40$.

2. A fast system will generate far more false signals.
3. The loss per trade in the slower system will be greater than the loss for the corresponding trade in the faster system. In some cases, a fast system might even realize a small profit on a minor trend that results in a loss in a slower system. For example, the $N = 40$ system's August buy signal that was liquidated in November resulted in a net loss of approximately \$2.54 (excluding commissions). The corresponding buy signal for the $N = 7$ version—triggered in July and exited in September—resulted in a net *gain* of around \$2.46.

As indicated by the preceding illustration, fast and slow systems will each work better under different circumstances. In the case of the chosen illustration, on balance, the slow system was much more successful. Of course, one could just as easily have chosen an example in which the reverse observation was true. However, empirical evidence suggests that, in most markets, slower systems tend to work better. In any case, the choice between a fast and a slow system must be based on up-to-date empirical testing.

The previous example of a breakout system was based on the current day's close and prior period's high and low. It should be noted that these choices were arbitrary. Other alternative combinations might include current day's high or low versus prior period's high or low; current day's close versus prior period's high close or low close; and current day's high or low versus prior period's high close or low close. Although the choice of the condition that defines a breakout will affect the results, the differences between the variations just given (for the same value of N) will be largely random and not overwhelming. Thus, while each of these definitions might be tested, it probably makes more sense to focus research efforts on more meaningful modifications of the basic system.

The pitfalls of breakout-type systems are basically the same as those of moving average systems and are detailed in the following section.

Ten Common Problems with Standard Trend-Following Systems

1. **Too many similar systems.** Many different trend-following systems will generate similar signals. Thus, it is not unusual for a number of trend-following systems to signal a trade during the same one- to five-day period. Because many speculators and futures funds base their decisions on basic trend-following systems, their common action can result in a flood of similar orders. Under such circumstances, traders using these systems may find their market and stop orders filled well beyond the intended price, if there is a paucity of offsetting orders.
2. **Whipsaws.** Trend-following systems will signal all major trends; the problem is that they will also generate many false signals. A major frustration experienced by traders using trend-following systems is that markets will frequently move far enough to trigger a signal and then reverse direction. This unpleasant event can even occur several times in succession; hence, the term *whipsaw.* For example, Figure 16.6, which indicates the trade signals generated by a breakout system (close beyond prior N-day high-low) for $N = 10$, provides a vivid illustration of the dark side of trend-following systems.

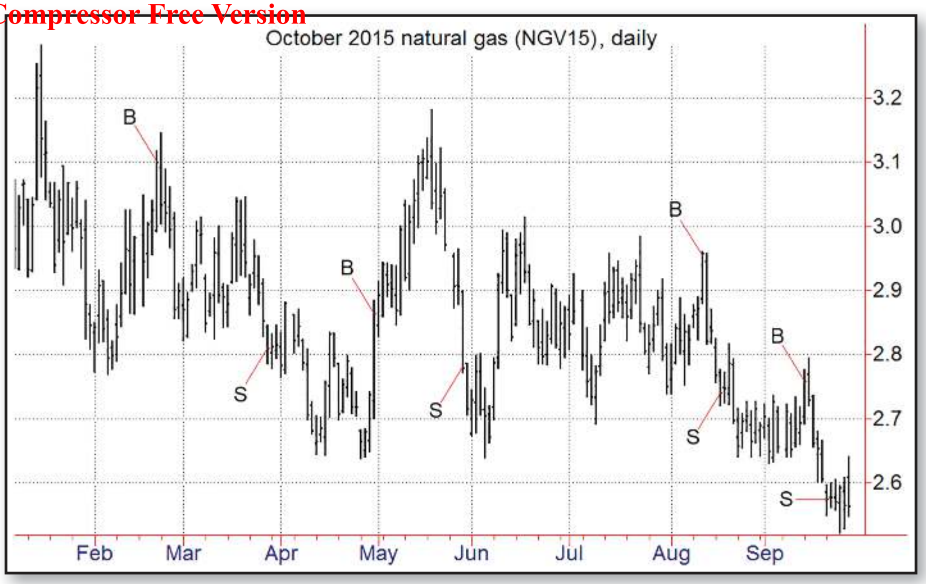

FIGURE 16.6 Breakout Signals in Trading Range Market: October 2015 Natural Gas Futures
Note: B = buy signal: close above prior 10-day high; S = sell signal: close below 10-day low.
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

3. **Failure to exploit major price moves.** Basic trend-following systems always assume an equal-unit-size position. As a result, given an extended trend, the best such a system can do is to indicate a one-unit position in the direction of the trend. For example, in Figure 16.7 a breakout system with $N = 40$ would signal a long position in December 2012 and remain long throughout the entire uptrend until February 2014. Although this outcome is hardly unfavorable, profitability could be enhanced if the trend-following system were able to take advantage of such extended trends by generating signals indicating increases in the base position size.
4. **Nonsensitive (slow) systems can surrender a large percentage of profits.** Although slow variations of trend-following systems may often work best, one disturbing feature of such systems is that they may sometimes surrender a large portion of open profits. In Figure 16.8, for example, a breakout system with $N = 40$ catches a major portion of the October–December 2014 price advance in silver, but then surrenders more than the entire gain before an opposite signal occurs. The June buy signal is initially profitable, but then realizes a much larger loss by the time a sell signal is received.
5. **Cannot make money in trading range markets.** The best any trend-following system can do during a period of sideways price action is to break even—that is, generate no new trade signals. In most cases, however, trading range markets will be characterized by whipsaw losses. This is a particularly significant consideration since sideways price action represents the predominant state of most markets.
6. **Temporary large losses.** Even an excellent trend-following system may witness transitory periods of sharp equity retracement. Such events can be distressing to the trader who enjoys a profit cushion, but they can be disastrous to the trader who has just begun following the system's signals.

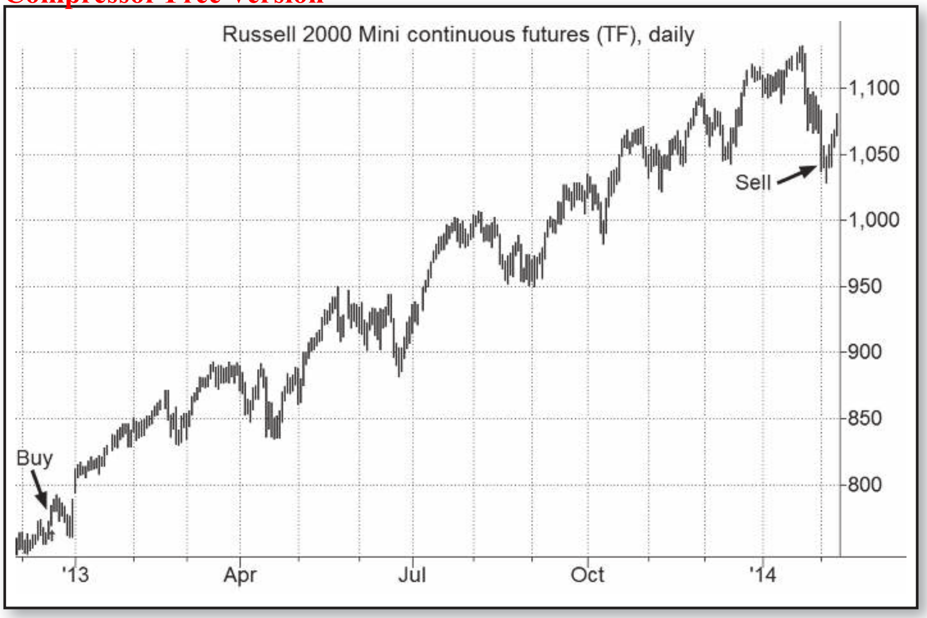

FIGURE 16.7 Failure of System to Exploit Major Price Move: Russell 2000 Mini Futures

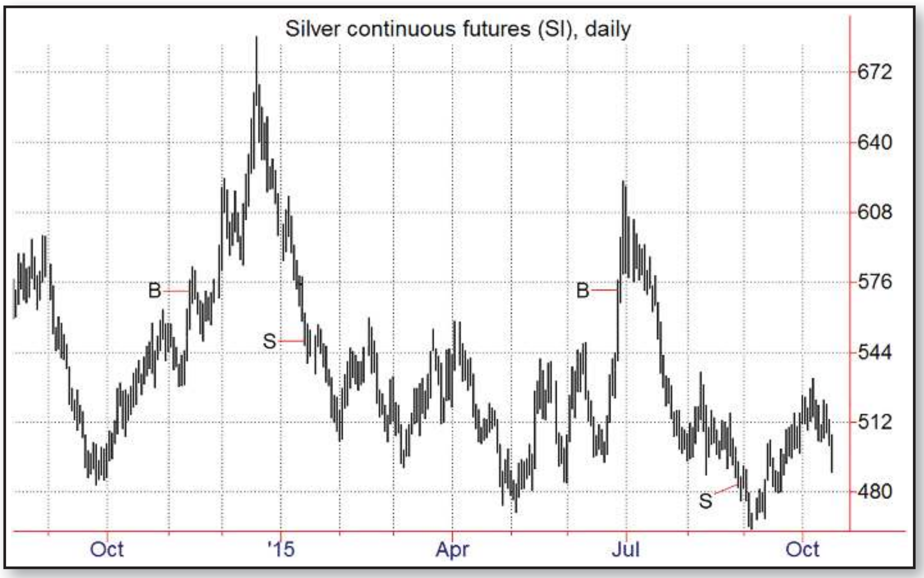

FIGURE 16.8 Surrender of Profits by Nonsensitive System: Silver Continuous Futures

Note: B = buy signal: close above prior 40-day high; S = sell signal: close below 40-day low.

7. **Extreme volatility in best-performing systems.** In some cases, the trader may find that the most profitable trend-following systems are also subject to particularly sharp retracements, thereby implying an unacceptable level of risk.
8. **System works well in testing but then bombs.** This scenario is perhaps the most common tale of woe among traders who have used mechanical trading systems.
9. **Parameter shift.**[2] Frequently, the trader may perform an exhaustive search to find the best variation of a system based on past data (e.g., the optimum value of N in a breakout system), only to find that the same variation performs poorly (relative to other variations) in the ensuing period.
10. **Slippage.** Another common experience: the system generates profits on paper, but *simultaneously* loses money in actual trading. Slippage is discussed in Chapter 19.

Possible Modifications for Basic Trend-Following Systems

Even simple systems, such as moving average or breakout systems, will probably prove profitable if traded consistently over a broad range of markets for a sufficient length of time (e.g., three to five years or longer). However, the simplicity of these systems is a vice as well as a virtue. In essence, the rules of these systems are perhaps too simple to adequately account for the wide variety of possible market situations. Even if net profitable over the long run, simple trend-following systems will typically leave the trader exposed to periodic sharp losses. In fact, the natural proclivity of many, if not most, users of such systems to abandon the approach during a losing period will lead them to experience a net loss even if the system proves profitable over the longer run.

In this section, we discuss some of the primary ways to modify basic trend-following systems in an effort to improve their performance. For simplicity, most of the examples will use the previously described simple breakout system. However, the same types of modifications could also be applied to other basic trend-following systems (e.g., crossover moving average).

Confirmation Conditions

An important modification that can be made to a basic trend-following system is the requirement for additional conditions to be met before a signal is accepted. If these conditions are not realized before an opposite direction signal is received, no trade occurs. Confirmation rules are designed specifically to deal with the nemesis of trend-following systems: false signals. The idea is that valid signals will fulfill the confirmation conditions, while false signals generally will not. The range of possible

[2] The meaning of the term *parameter* as it is used in trading systems is detailed in Chapter 19.

FIGURE 16.9 Penetration as Confirmation Condition: Coffee Continuous Futures
Note: B, S = signals for breakout system with $N = 12$; B, S = signals for breakout system with $N = 12$ and 3 percent closing penetration confirmation.

choices for confirmation conditions is limited only by the imagination of the system designer. Here are three examples:

1. **Penetration.** A trade signal is accepted only if the market moves a specified minimum amount beyond a given reference level (e.g., signal price). This confirming price move can be measured in either nominal or percentage terms. Figure 16.9 compares the trade signals generated by a standard breakout system with $N = 12$ and the corresponding system with a confirmation rule requiring a close that exceeds the prior N-day high or low by at least 3 percent.[3] Note that in this example, although the confirmation rule results in moderately worse entry levels for valid signals, it eliminates five of six losing buy signals. (The sell signals following the nonconfirmed buy signals are also eliminated, since the system is already short at these points.)
2. **Time delay.** In this approach, a specified time delay is required, at the end of which the signal is reevaluated. For example, a confirmation rule may specify that a trade signal is taken if the market closes beyond the signal price (higher for a buy, lower for a sell) at any time six or more days beyond the original signal date. Figure 16.10 compares the signals generated by a basic breakout system with $N = 12$, and the corresponding system with the six-day time delay confirmation condition. Again, the confirmation rule eliminates five of the six losing buy signals.

[3] Because Figure 16.9 depicts a continuous futures series, percentage price changes would be equal to the price changes shown on this chart divided by the corresponding nearest futures price, which is not shown. Recall from Chapter 5 that continuous futures accurately reflect price *swings* but not price *levels.* Consequently, continuous futures cannot be used as the divisor to calculate percentage changes.

FIGURE 16.10 Time Delay as a Confirmation Condition: Coffee Continuous Futures
Note: B, S = signals for breakout system with $N = 12$; Ⓑ, Ⓢ = signals for breakout system with $N = 12$ and six-day time delay confirmation.

3. **Pattern.** This is a catch-all term for a wide variety of confirmation rules. In this approach, a specified pattern is required to validate the basic system signal. For example, the confirmation rule might require three subsequent thrust days beyond the signal price.[4] Figure 16.11 compares the signals generated by the basic breakout system, with $N = 12$ and the signals based upon the corresponding system using the three-thrust-day validation condition. The thrust-day count at confirmed signals is indicated by the numbers on the chart. Here, too, the confirmation rule eliminates five of six losing buy signals.

The design of trading systems is a matter of constant trade-offs. The advantage of confirmation conditions is that they will greatly reduce whipsaw losses. However, it should be noted that confirmation rules also have an undesirable side effect—they will delay entry on valid signals, thereby reducing gains on profitable trades. For example, in Figures 16.9 through 16.11, note that the confirmation rules result in worse entry prices for all the valid trade signals. The confirmation condition will be beneficial as long as reduced profits due to delayed entry are more than offset by avoided losses. A system that includes confirmation conditions will not always outperform its basic system counterpart, but if properly designed it will perform significantly better over the long run.

[4] A thrust day, which was originally defined in Chapter 9, is a day with a close above the previous day's high or below the previous day's low.

FIGURE 16.11 Example of a Pattern Confirmation Condition: Coffee Continuous Futures
Note: B, S = signals for breakout system with $N = 12$; B, S = signals for breakout system with $N = 12$ and three-thrust-day confirmation.

Filter

The purpose of a filter is to eliminate those trades that are deemed to have a lower probability of success. For example, the technical system might be combined with a fundamental model that classifies the market as bullish, bearish, or neutral. Technical signals would then be accepted only if they were in agreement with the fundamental model's market designation. In cases of disagreement, a neutral position would be indicated. In most cases, however, the filter condition(s) will also be technical in nature. For example, if one could derive a set of rules that had some accuracy in defining the presence of a trading range market, signals that were received when a trading range market was indicated would not be accepted. In essence, in developing a filter, the system designer is trying to find a common denominator applicable to the majority of losing trades.

We will use the frequently unsatisfactory simple moving average system to provide a specific example of a filter condition. The noncircled signals in Figure 16.1 illustrate the typical tendency of the simple moving average system to generate many false signals—even in trending markets. These whipsaw trades can be substantially reduced by applying a filter rule that requires trade signals to be consistent with the trend of the moving average. For example, price crossing the moving average from below and closing above the moving average would be accepted as a buy signal only if the moving average was up relative to the previous day's level. This filter condition makes intuitive sense because it adheres to the basic technical concept of trading with the major trend.

Two points should be clarified regarding the application of this rule:

1. A rejected signal could be activated later if the moving average subsequently turned in the direction of the signal *before* an opposite-direction crossover of the price and moving average.
2. Signals that occur after rejected signals are ignored because the net position is already consistent with the implied trade. This observation is true because the simple moving average system is always in the market.

The circled signals in Figure 16.1 indicate the trades that would have been accepted if the filter rule just described were applied. (In both instances these trades occurred after delays, as previously described, rather than upon immediate penetration of the moving average.) As can be seen, the rule substantially reduces the number of false signals. Although in some cases the application of the filter condition results in adversely delayed trade entries (for example, the July sell signal), on balance the benefits clearly outweigh the disadvantages. Of course, a single illustration doesn't prove anything. However, the implication of Figure 16.1 does have a more general applicability. Most empirical testing would reveal that, more often than not, the inclusion of the type of filter rule depicted in Figure 16.1 tends to improve performance.

In fact, a crossover between price and the moving average that is opposite to the direction of the moving average trend can often provide a good signal to *add to rather than reverse* the original position. For example, in Figure 16.1 the March and May 2014 downside penetrations of the moving average could be viewed as buy rather than sell signals because the moving average trend was still up in those instances. The rationale behind this interpretation is that in a trending market, reactions often carry to the vicinity of a moving average before prices resume their longer-term trend (see Chapter 12). Thus, in effect, such rejected signals could actually provide the basis for a method of pyramiding.

It should be noted that, in a sense, the confirmation conditions detailed in the previous section represent one type of filter, insofar as signals that fulfill a subsequent set of conditions are accepted, while those that do not are eliminated. However, the distinction here is that a filter implies a set of screening rules applied *at the time* the base system signal is received. In other words, the sorting procedure occurs without any dependency on subsequent developments (although, to be perfectly accurate, subsequent developments could still permit a delayed acceptance of a rejected signal). Consequently, as we have defined the terms, a system can include both a filter and a confirmation rule. In such a system, only signals that were accepted based on the filter definition and subsequently validated by the confirmation rule(s) would actually result in trades.

Market Characteristic Adjustments

One criticism of simple trend-following systems is that they treat all markets alike. For example, in a breakout system, with $N = 20$, both highly volatile and very quiet markets will require the same conditions for a buy signal—a 20-day high. Market characteristic adjustments seek to compensate for the fact that a system's optimum parameter value settings will depend on market conditions. For example, in the case of a breakout system, instead of using a constant value for N, the relevant value for N might be contingent on the market's volatility classification. As a specific illustration, the average two-day price range during the past 50-day period might be used to place the market into one of five volatility classifications.[5] The value of N used to generate signals on any given day would then depend on the prevailing volatility classification.

Volatility appears to be the most logical choice for classifying market states, although other criteria could also be tested (e.g., fundamentally based conditions, average volume level). In essence, this type of modification seeks to transform a basic trend-following system from a static to a dynamic trading method.

Differentiation between Buy and Sell Signals

Basic trend-following systems typically assume analogous conditions for buy and sell signals (e.g., buy on close above 20-day high, sell on close below 20-day low). However, there is no reason to make this assumption automatically. It can be argued that bull and bear markets behave differently. For example, a survey of a broad spectrum of historical price data would reveal that price breaks from major tops tend to be more rapid than price rallies from major bottoms.[6] This observation suggests a rationale for using more sensitive conditions to generate sell signals than those used to generate buy signals. However, the system designer using such an approach should be particularly sensitive to the danger of overfitting the system—a pitfall discussed in detail in Chapter 19.

Pyramiding

One inherent weakness in basic trend-following systems is that they automatically assume a constant unit position size for all market conditions. It would seem desirable to allow for the possibility of larger position sizes in the case of major trends, which are almost entirely responsible for the success of any trend-following system. One reasonable approach for adding units to a base position in a major trend is to wait for a specified reaction and then initiate the additional unit(s) on evidence of a resumption of the trend. Such an approach seeks to optimize the timing of pyramid units, as well as to provide exit rules that reasonably limit the potential losses that could be incurred by such added positions. An

---

[5] A two-day price range is used as a volatility measure instead of a one-day range since the latter can easily yield a distorted image of true market volatility. For example, on a limit day, the one-day range would equal zero, in extreme contrast to the fact that limit days reflect highly volatile conditions. Of course, many other measures could be used to define volatility.

[6] The reverse statement would apply to short-term interest rate markets, which are quoted in terms of the instrument price, a value that varies inversely with the interest rate level. In the interest rate markets, interest rates rather than instrument prices are analogous to prices in standard markets. For example, there is no upper limit to a commodity's price or interest rates, but the downside for both of these items is theoretically limited. As another example, commodity markets tend to be more volatile when prices are high, while short-term interest rate markets tend to be more volatile when interest rates are high (instrument prices are low). The situation for long-term (i.e., bond) markets is ambiguous since although interest rates can fall no lower than approximately zero, the pricing mathematics underlying these instruments result in an accelerated price advance (for equal interest rate changes) as interest rates fall.

example of this type of approach was detailed in Chapter 12. Another example of a possible pyramid strategy would be provided by the following set of rules:

Buy Case

1. A reaction is defined when the net position is long and the market closes below the prior 10-day low.
2. Once a reaction is defined, an additional long position is initiated on any subsequent 10-day high if the following conditions are met:
   a. The pyramid signal price is above the price at which the most recent long position was initiated.
   b. The net position size is less than three units. (This condition implies that there is a limit of two pyramid units.)

Sell Case

1. A reaction is defined when the net position is short and the market closes above the prior 10-day high.
2. Once a reaction is defined, an additional short position is initiated on any subsequent 10-day low if the following conditions are met:
   a. The pyramid signal price is below the price at which the most recent short position was initiated.
   b. The net position size is less than three units. (This condition implies that there is a limit of two pyramid units.)

Figure 16.12 illustrates the addition of this pyramid plan to a breakout system with $N = 40$ applied to the 2012–2013 gold market. (For now, ignore the "stop level" signals; they are explained shortly.)

Risk control becomes especially important if a pyramiding component is added to a system. Generally speaking, it is usually advisable to use a more sensitive condition for liquidating a pyramid position than the condition required to generate an opposite signal. The following is one example of a set of stop rules that might be employed in a system that uses pyramiding. Liquidate all pyramid positions whenever either condition is fulfilled:

1. An opposite trend-following signal is received.
2. The market closes above (below) the high (low) price since the most recently defined reaction that was followed by a pyramid sell (buy). Figure 16.12 illustrates the stop levels implied by this rule in the case of the 2012–2013 gold market.

Trade Exit

The existence of a trade exit rule in a system (e.g., a stop rule) would permit the liquidation of a position prior to receiving an opposite trend-following signal. Such a rule would serve to limit losses

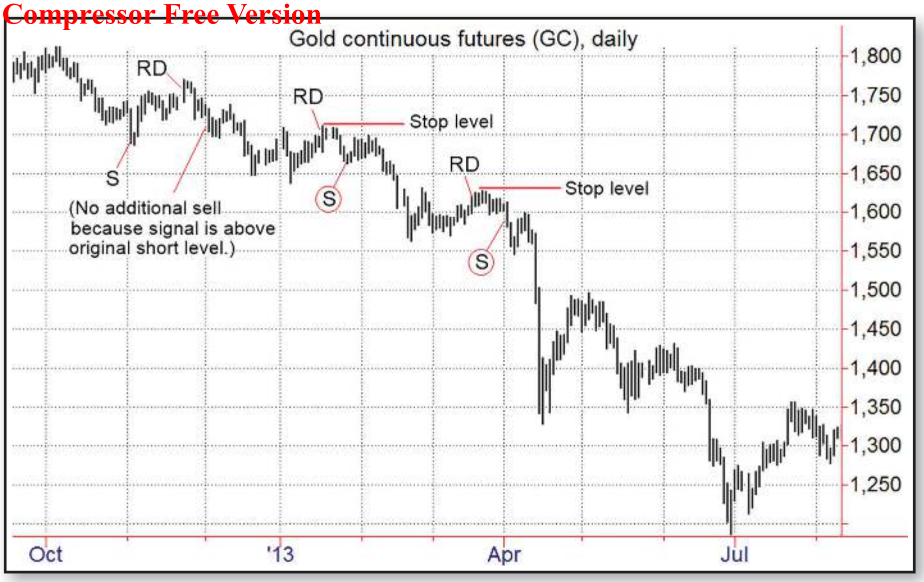

FIGURE 16.12 Pyramid Signals: Gold Continuous Futures
Note: S = base position sell signal; (S) = pyramid sell signal; RD = reaction defined.

on losing trades as well as limit the amount of open profits surrendered on winning trades. Although these are highly desirable goals, the trade-off implied by using a trade exit rule is relatively severe. If a trade exit rule is used, rules must be specified for reentering the position; otherwise, the system will be vulnerable to missing major trends.

The danger in using a trade exit rule is that it may result in the premature liquidation of a good trade. Although the reentry rule will serve as a backstop, the combination of an activated trade exit rule and a subsequent reentry is a whipsaw loss. Thus, it will not be at all uncommon for the addition of a trade exit rule (and implied reentry rule) to have a negative impact on performance. Nevertheless, although it is not easy, for some systems it will be possible to structure trade exit rules that improve performance on balance. (In terms of return, and usually in terms of return/risk measures as well, if a trade exit rule helps performance, the use of the trade exit rule as a reversal signal—as opposed to just a liquidation signal—will help performance even more.) Trade exit rules can also be made dynamic. For example, the trade exit condition can be made increasingly sensitive as a price move becomes more extended in either magnitude or duration.

Countertrend Systems

General Considerations Regarding Countertrend Systems

Countertrend systems often appeal to many traders because their ultimate goal is to buy low and sell high. Unfortunately, the difficulty of achieving this goal is inversely proportional to its desirability.

A critical distinction to keep in mind is that whereas a trend-following system is basically self-correcting, a countertrend system implies unlimited losses. Therefore, it is essential to include some stop-loss conditions in any countertrend system (unless it is traded simultaneously with trend-following systems). Otherwise, the system could end up being long for the duration of a major downtrend or short for the duration of a major uptrend. (Stop-loss conditions are optional for most trend-following systems, since an opposite signal will usually be received before the loss on a position becomes extreme.[7])

One important advantage of using a countertrend system is that it provides the opportunity for excellent diversification with simultaneously employed trend-following systems. In this regard, it should be noted that a countertrend system might be desirable even if it was a modest net loser, the reason being that if the countertrend system was inversely correlated to a simultaneously traded trend-following system, trading both systems might imply less risk than trading the trend-following system alone. Therefore, it is entirely possible that the two systems combined might yield a higher percent return (at the same risk level), even if the countertrend system alone lost money.

Types of Countertrend Systems

The following are some types of approaches that can be used to try to construct a countertrend system:

Fading minimum move. This is perhaps the most straightforward countertrend approach. A sell signal is indicated each time the market rallies by a certain minimum amount above the low point since the last countertrend buy signal. Similarly, a buy signal is indicated whenever the market declines by a minimum amount below the high point since the last countertrend sell signal. The magnitude of the price move required to generate a trade signal can be expressed in either nominal or percentage terms. Figure 16.13 illustrates the trade signals that would be generated by this type of countertrend system for a 7.5 percent threshold level in the January–September 2015 natural gas market. It is no accident this chart depicts the same market that was previously used in this chapter to illustrate whipsaw losses for a sensitive trend-following system (see Figure 16.6). Countertrend systems will tend to work best under those types of market conditions in which trend-following systems fare poorly.

Fading minimum move with confirmation delay. This is similar to the preceding countertrend system, with the exception that some minimum indication of a trend reversal is required before the countertrend trade is initiated. For example, a one-thrust-day confirmation might be required to validate countertrend signals based on fading a given percent price move.

Oscillators. A countertrend system could use oscillators to generate trade signals. However, as discussed in Chapters 11 and 12, although using oscillators to signal countertrend trades may work well in a trading-range market, in a trending market such an approach can be disastrous.

[7] Stop-loss rules, however, might be mandatory for an extremely nonsensitive trend-following system—for example, a breakout system with $N = 150$.

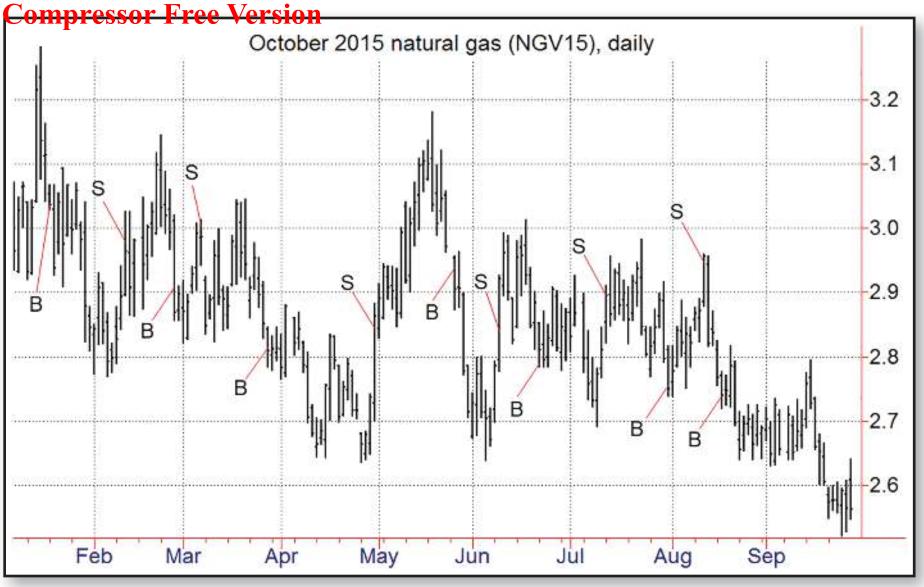

FIGURE 16.13 Countertrend Signals: October 2015 Natural Gas Futures

Note: Percentages are calculated as price changes in continuous futures divided by corresponding *nearest* futures price levels. B = buy signal: 7.5% decline from prior high; S = sell signal: 7.5% advance from prior low.

Contrary opinion. A countertrend system might use contrary opinion as an input in timing trades. For example, once the contrary opinion rose above a specified level, a short position would be indicated contingent on confirmation by a very sensitive technical indicator. (Contrary opinion was discussed in Chapter 14.)

Diversification

The standard interpretation attached to the term *diversification* is that trading is spread across a broad range of markets. Although this is the single most important type of diversification, assuming the availability of sufficient funds, there are two additional levels of possible diversification. First, each market can be traded with several systems. Second, several variations of each system can be used. For example, if two contracts of cocoa are being traded using the breakout system, each contract can be traded using a different value of N (i.e., the number of days whose high or low must be penetrated to trigger a signal).

In the following discussion, the term *single market system variation (SMSV)* will refer to the concept of a specific variation of a given system traded in a single market. Thus, the simple breakout system, with $N = 20$, traded in the cocoa market would be an example of an SMSV. In the simplest case in which a single system is used for all markets, and a single system variation is used in each market, there would be only one SMSV for each market traded. This simplified case represents the typical application of trading systems and employs only the standard diversification across markets. However, if sufficient funds are available, additional benefits can be obtained by also diversifying across different systems and different variations of each system.

There are three important benefits to diversification:

1. **Dampened equity retracements.** Different SMSVs will not witness their losses at precisely the same periods. Thus, by trading a wide variety of SMSVs, the trader can achieve a smoother equity curve. This observation implies that trading 10 SMSVs with equivalent profit/risk characteristics could provide lower risk at the same return level than trading 10 units of a single SMSV. Or, alternatively, by trading larger size, 10 SMSVs with equivalent profit/risk characteristics could provide higher return at the same risk level than trading 10 units of a single SMSV. Up to a point, diversification would be beneficial even if the portfolio included SMSVs with poorer expected performance. A key consideration would be a given SMSV's correlation with the other SMSVs in the portfolio.
2. **Ensured participation in major trends.** Typically, only a few of the actively traded futures markets will witness substantial price trends in any given year. Because the majority of trades in most trend-following systems will lose money,[8] it is essential that the trader participate in the large-profit trades—that is, major trends. This is a key reason for the importance of diversification across markets.
3. **Bad luck insurance.** Futures systems trading, like baseball, is a game of inches. Given the right combination of circumstances, even a minute difference in the price movement on a single day could have an extraordinary impact on the profitability of a specific SMSV. To illustrate this point, we consider a breakout system ($N = 20$) with a confirmation rule requiring a single thrust day that penetrates the previous day's high or low by a minimum amount. In system A this amount is 0.05 cents; in system B it is 0.10 cents. This is the only difference between the two systems.

Figure 16.14 compares these two systems for the December 1981 coffee market and represents the most striking instance I have ever encountered of the sensitivity of system performance to minute changes in system values. The basic system buy signal (i.e., close above the 20-day high) was received on July 16. This buy was confirmed by system A on July 17 as the close was 0.09 cents above the previous day's high (point A1). System B, however, which required a 0.10-cent penetration, did not confirm the signal until the following day (point B1).

The buy signal for system A would have been executed at approximately \$0.97 (point A2). However, due to the ensuing string of limit moves, the buy signal for system B could not be filled until prices surpassed \$1.22 (point B2). Thus, during this short interim, system A gained

[8] Such systems can still be profitable because the average gain significantly exceeds the average loss.

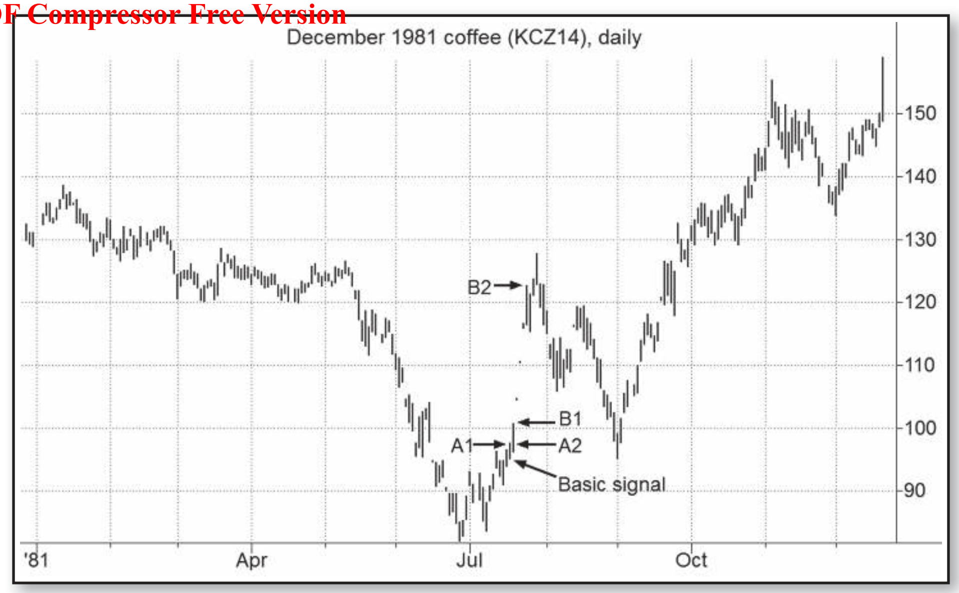

FIGURE 16.14 System Trading: A Game of Inches (December 1981 Coffee)

25¢/lb ($9,375 per contract), while system B, which was unable to reverse its short position, lost a similar amount. The failure of the market to close 0.01 cent higher on a given day (a price move equivalent to less than $4) resulted in an incredible $18,750 per contract difference in the performance of the two nearly identical system variations! It should be emphasized this example reflects the randomness in commodity price movements rather than the instability of the tested system. Any system, other than a day trading system, could reflect the same degree of instability, since the performance difference was due to just a single trade in which the signals were separated by only one day.

This example should explain how it is possible for a trader to lose money in a given market using a system that generally performs well—he may just have chosen a specific variation that does much worse than most other variations (even very similar ones). By trading several variations of a system, the speculator could mitigate the impact of such isolated, abnormally poor results.[9] Of course, in so doing, the trader would also eliminate the possibility of gains far exceeding the average performance of the system. On balance, however, this prospect represents a desirable trade-off, since it is assumed that the basic trading goal is consistent performance rather than windfall profits.

---

[9] In the preceding example, system A and system B were deliberately chosen to be nearly identical in order to make the point about the potential impact of chance in its strongest possible form. However, in practice, the trader should choose system variations that are substantially more differentiated.

Ten Common Problems with Trend-Following Systems Revisited

We are now ready to consider possible solutions to the previously enumerated problems with standard trend-following systems. The problems and the possible solutions are summarized in Table 16.2.

TABLE 16.2 Problems with Standard Trend-Following Systems and Possible Solutions

	Problems with Standard Trend-Following Systems		Possible Solutions
1.	Too many similar systems	1a.	Try to construct original systems in order to avoid the problem of "trading with the crowd."
		1b.	If trading more than one contract, spread out entry.
2.	Whipsaws	2a.	Employ confirmation conditions.
		2b.	Develop filter rules.
		2c.	Employ diversification.
3.	Failure to exploit major price moves	3.	Add pyramiding component.
4.	Nonsensitive (slow) systems can surrender a large percentage of profits.	4.	Employ trade exit rules.
5.	Cannot make money in trading-range markets	5.	Trade trend-following systems in conjunction with countertrend systems.
6.	Temporary large losses	6a.	If funds permit, trade more than one system in each market.
		6b.	When beginning to trade a system, trade more lightly if entering positions at a point after the signal has been received.
7.	Extreme volatility in best-performing systems	7.	By employing diversification, the trader can allocate some funds to a high-profit-potential system that is too risky to trade on its own.
8.	System works well in testing but then bombs.	8.	The danger of such a development can be reduced if systems are properly tested. This subject is discussed in detail in Chapter 19.
9.	Parameter shift	9a.	If funds permit, diversify by trading several variations of each system.
		9b.	Experiment with systems that incorporate market characteristic adjustments.
10.	Slippage	10.	Use realistic assumptions (discussed in Chapter 19).

CHAPTER 17

Examples of Original Trading Systems

Nothing works at all times in all kinds of markets.

—Adam Smith

The previous chapter provided two examples of generic trading systems—moving averages and breakouts. This chapter details several original trading systems that are based on some of the patterns introduced in Chapter 9. Although the systems detailed here represent fully automated trading strategies, the primary goal of this chapter is not to offer specific trading systems, but rather to give readers a feel for how technical concepts can be utilized to construct a mechanical trading approach. Studying these examples should provide readers with ideas as to how to design their own trading systems.

Wide-Ranging-Day System

Basic Concept

A wide-ranging day, which was introduced in Chapter 9, is a day with a much wider *true range*[1] than recent trading sessions. The high volatility inherent in wide-ranging days gives these days special significance. Typically, the market will tend to extend in the direction of the initial price move beyond

[1] The *true range* is equal to the *true high* minus the *true low*. The *true high* is the maximum of the current day's high and the previous day's close. The *true low* is the minimum of the current day's low and the previous day's close.

the boundaries of the wide-ranging day. However, situations in which the market originally penetrates one side of the wide-ranging day and then reverses to penetrate the other side also have significance.

The wide-ranging-day system defines trading ranges based on wide-ranging days. Signals are generated when prices close above or below these trading ranges. In the simplest case, the trading range is defined as the wide-ranging day itself. However, we make the system more general by defining the trading range as the price range encompassing all the true highs and true lows during the period extending from $N1$ days before the wide-ranging day to $N2$ days after, where $N1$ and $N2$ are parameter values that must be defined. For example, if both $N1$ and $N2$ equal 0, the trading range would be defined by the wide-ranging day itself (i.e., the range between the true high and true low of the wide-ranging day). If $N1 = 2$ and $N2 = 4$, the trading range would be defined as the range between the highest true high and lowest true low in the interval beginning two days before the wide-ranging day and ending four days after it.

Definitions

Wide-ranging day. A day on which the *volatility ratio* (VR) is greater than k (e.g., $k = 2$). The VR is equal to today's true range divided by the average true range of the past N-day period (e.g., $N = 10$).

Price trigger range (PTR). The range defined by the highest true high and lowest true low in the interval between $N1$ days before the most recent wide-ranging day to $N2$ days after. Note that the PTR cannot be defined until $N2$ days after a wide-ranging day. (If $N2 = 0$, the PTR would be defined as of the close of the wide-ranging day itself.) The PTR will be redefined each time there is a new wide-ranging day (i.e., $N2$ days after such an event).

Trading Signals

Buy case. On a close above the high of the PTR, reverse from short to long.

Sell case. On a close below the low of the PTR, reverse from long to short.

Daily Checklist

To generate trading signals, perform the following steps each day:

1. If short and today's close is above the high of the PTR, liquidate short and go long.
2. If long and today's close is below the low of the PTR, liquidate long and go short.
3. Check whether exactly $N2$ days have elapsed since the most recent wide-ranging day. If this condition is met, redefine the PTR.

The order of these steps is very important. Note that the check for new trading signals *precedes* the check whether the PTR should be redefined. Thus, if the day a new PTR is defined also signals a trade based on the prevailing PTR going into that day, a signal would be generated. If step 3 preceded steps 1 and 2, trade signals could get delayed each time a signal occurred on the day a new PTR is defined ($N2$ days after the most recent wide-ranging day, which would be the wide-ranging day itself when $N2 = 0$). For example, assume the system is long, $N2 = 0$, and the close on a new wide-ranging day is below the low of the preceding wide-ranging day. According to the listed step order, the new wide-ranging day would signal a reversal from long to short. If steps 1 and 2 followed step 3, no signal would occur, since the PTR would be redefined, and the market would have to close below the *new* wide-ranging day to trigger a signal.

System Parameters

***N*1.** The number of days prior to the wide-ranging day included in the PTR period.

***N*2.** The number of days after the wide-ranging day included in the PTR period.

***k*.** The value the volatility ratio (VR) must exceed in order to define a wide-ranging day.

Note: N, the number of past days used to calculate the VR, is assumed to be fixed (e.g., N = 10).

Parameter Set List

Table 17.1 provides a sample parameter set list. Readers can use this list as is or adjust it as desired. The subject of testing multiple parameter sets and deciding which one to use in actual trading is addressed in Chapter 19.

TABLE 17.1 Parameter Set List

	k	*N1*	*N2*
1.	1.6	0	0
2.	1.6	2	0
3.	1.6	4	0
4.	1.6	0	2
5.	1.6	2	2
6.	1.6	4	2
7.	1.6	0	4
8.	1.6	2	4
9.	1.6	4	4
10.	2.0	0	0
11.	2.0	2	0
12.	2.0	4	0
13.	2.0	0	2
14.	2.0	2	2
15.	2.0	4	2
16.	2.0	0	4
17.	2.0	2	4
18.	2.0	4	4
19.	2.4	0	0
20.	2.4	2	0
21.	2.4	4	0
22.	2.4	0	2
23.	2.4	2	2
24.	2.4	4	2
25.	2.4	0	4
26.	2.4	2	4
27.	2.4	4	4

An Illustrated Example

To illustrate how the system works, Figures 17.1 through 17.5 superimpose trading signals on copper charts spanning late October 2013 to November 2015, a period the weekly chart inset in Figure 17.1 shows consisted mostly of a choppy, longer-term price descent, interspersed with short-term uptrends in mid-2014 and early 2015. Note these charts are continuous futures to coincide with the price series used to generate signals. As will be fully detailed in the next two chapters, continuous futures are usually the most suitable price series to use in trading systems. To help provide continuity between charts, each chart overlaps one to two months of the preceding chart.

Two types of signals are indicated on the accompanying charts:

1. The noncircled signals are generated by the system when both *N*1 and *N*2 are set to zero. In other words, the PTR is defined by the true high and true low of the wide-ranging day.
2. The circled signals are generated by the system when *N*1 = 2 and *N*2 = 4. (In other words, the PTR is defined by the true price range encompassing the interval beginning two days before the wide-ranging day and ending four days after it.)

Occasionally, both sets of parameter values will yield identical signals. In most cases, however, the second system version will trigger signals later or not at all. (The reverse can never occur, since the PTR based on *N*1 = 2 and *N*2 = 4 must be at least as wide as the PTR based on *N*1 = 0 and *N*2 = 0. Therefore any penetration of the former PTR must also be a penetration of the latter PTR, but not vice versa.)

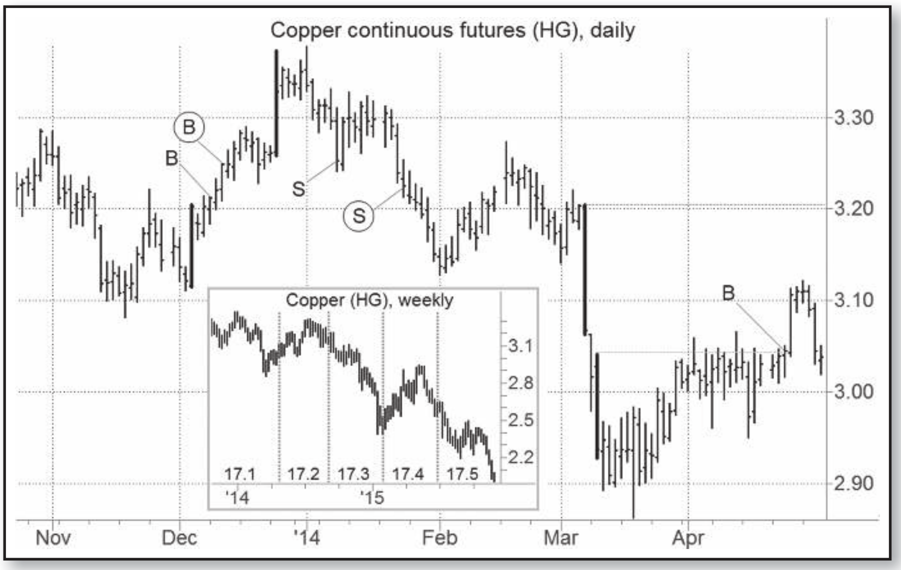

FIGURE 17.1 Wide-Ranging Day System, Chart 1: Copper Continuous Futures
Note: Thicker bars are wide-ranging days. B, S = buy and sell signals for *N*1 = 0 and *N*2 = 0; (B), (S) = buy and sell signals for *N*1 = 2 and *N*2 = 4.

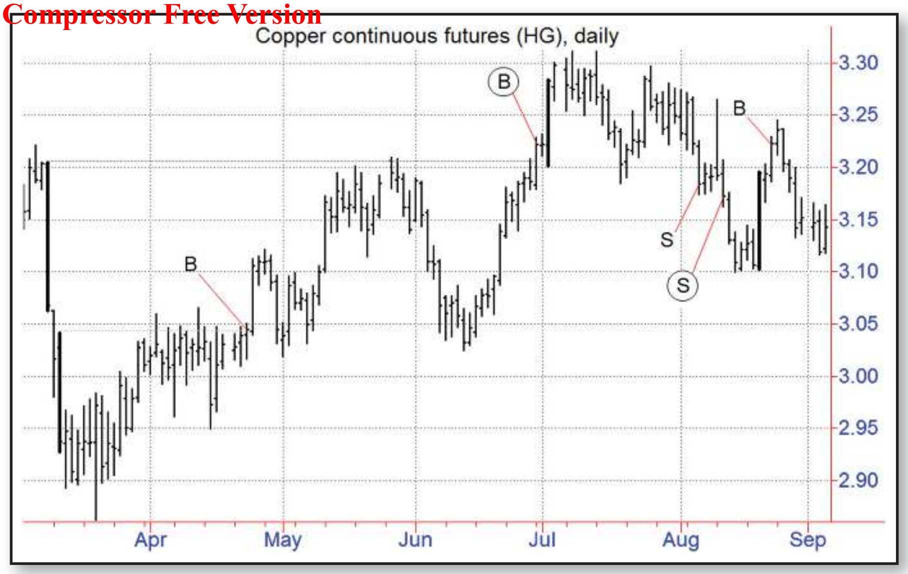

FIGURE 17.2 Wide-Ranging Day System, Chart 2: Copper Continuous Futures
Note: Thicker bars are wide-ranging days. B, S = buy and sell signals for $N1 = 0$ and $N2 = 0$; Ⓑ, Ⓢ = buy and sell signals for $N1 = 2$ and $N2 = 4$.

FIGURE 17.3 Wide-Ranging Day System, Chart 3: Copper Continuous Futures
Note: Thicker bars are wide-ranging days. B, S = buy and sell signals for $N1 = 0$ and $N2 = 0$; Ⓑ, Ⓢ = buy and sell signals for $N1 = 2$ and $N2 = 4$.

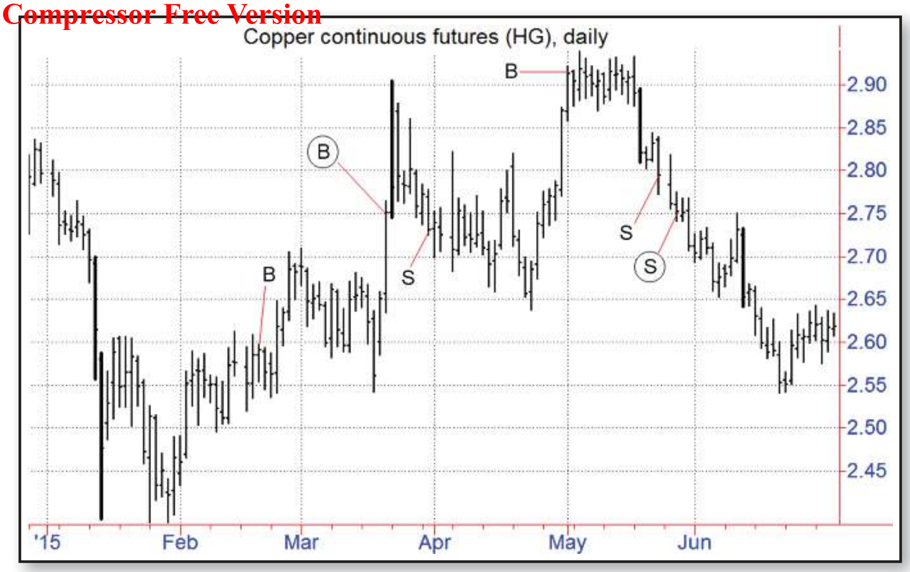

FIGURE 17.4 Wide-Ranging Day System, Chart 4: Copper Continuous Futures
Note: Thicker bars are wide-ranging days. B, S = buy and sell signals for $N1 = 0$ and $N2 = 0$; (B), (S) = buy and sell signals for $N1 = 2$ and $N2 = 4$.

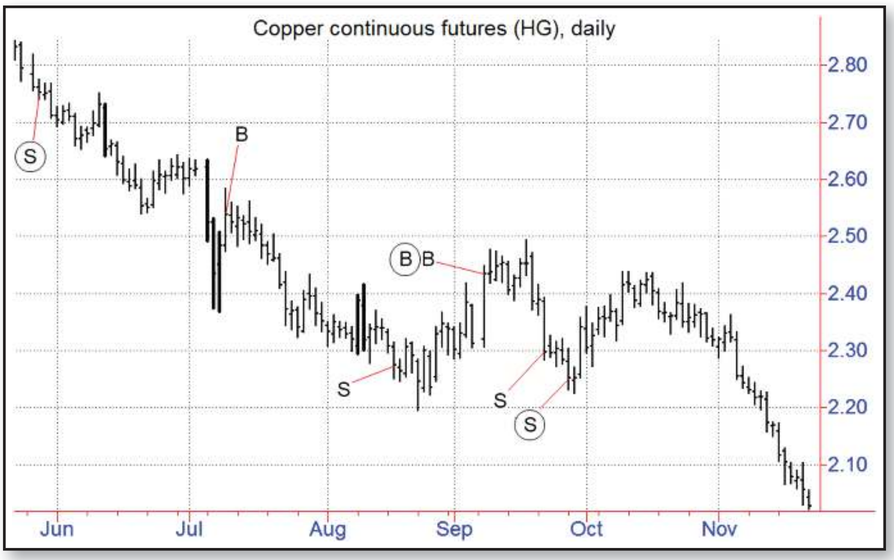

FIGURE 17.5 Wide-Ranging Day System, Chart 5: Copper Continuous Futures
Note: Thicker bars are wide-ranging days. B, S = buy and sell signals for $N1 = 0$ and $N2 = 0$; (B), (S) = buy and sell signals for $N1 = 2$ and $N2 = 4$.

First, we examine the trading signals generated for the system version when both $N1$ and $N2$ equal zero (the noncircled signals). Therefore, for now, ignore the circled signals, which are based on the parameter set consisting of $N1 = 2$ and $N2 = 4$. We will subsequently examine the instances in which the two parameter sets yield different signals.

The first signal occurs in December 2013 when a close above the high of the December 4 wide-ranging day triggers a buy (Figure 17.1). The system then reverses to short modestly higher in January 2014 when the market closes below the low of the wide-ranging day formed in late December. The January short position profits from the ensuing downtrend and remains intact until late April when the market closes above the high of the second wide-ranging day that formed in March, triggering a buy signal.

The April 2014 long position remains intact for several months, capturing a portion of the ensuing uptrend, until it is reversed in early August when the market closes below the low of the early July wide-ranging day (Figure 17.2). The early August short position is short-lived and results in the first losing trade when a subsequent market bounce forms a wide-ranging day that is exceeded by a closing price two days later, triggering a buy signal. This August buy signal proves to be a whipsaw trade as the system reverses back to short in September 2014 (Figure 17.3).

The next buy signal materializes near the same level in late October 2014 when the market closes above the true high of the October wide-ranging day. This buy signal proves to be another whipsaw loss as the market immediately turns lower and eventually closes below the same October wide-ranging day, triggering a sell signal in November. Note, as is the case here, a single wide-ranging day can trigger multiple trades (in opposite directions) in the absence of intervening wide-ranging days. The November sell signal yields a small profit before leading to a third successive losing buy signal in December 2014. The January 2015 sell signal is exited at a profit in February 2015 when the market closes above the high of the second wide-ranging day formed in January (Figure 17.4). Additional trades are shown in Figures 17.4 and 17.5.

Next we examine how the signals generated by the second parameter set ($N1 = 2$, $N2 = 4$; circled on charts) differ from those that result from the first parameter set ($N1 = 0$, $N2 = 0$). One pattern the reader will notice is that whenever both parameter sets had signals in the same cycle—a signal in the same direction before the first parameter set triggered an opposite signal—the delay caused by using the second parameter set almost invariably resulted in a less favorable entry level. In most cases the differences in entry levels were moderate (e.g., the signals shown in Figure 17.1). In some instances, however, the difference in entry levels was quite substantial. For example, in Figure 17.2, the second parameter set went long in late June, more than two months after the first parameter set, because prices needed to close not just above the high of the March 11 wide-ranging day, but above the high of the two days preceding that day. Occasionally, both parameter sets may trigger signals on the same day (e.g., the September 2015 buy in Figure 17.5), but there are no instances where the second parameter set has a better entry. The poorer entry levels generated by the second parameter set are no accident, since the wider PTRs defined by the nonzero $N1$ and $N2$ values will always result in equal or higher buy signals and equal or lower sell signals.

The reader might well wonder why one would ever want to use nonzero values for *N*1 and *N*2, since the resulting delayed entries are invariably equal to or worse than entries based on keeping *N*1 and *N*2 equal to zero. The answer lies in the fact that the broader PTRs that result from nonzero *N*1 and *N*2 values will tend to filter out some losing signals—a characteristic that can have a major impact on the system's profitability. For example, following the August 2014 sell signal, the second parameter set avoids the three successive losing buy signals generated by the first parameter set (Figure 17.3). As a result, the second parameter set generates a substantial profit during this period while the series of trades generated by the first parameter set results in a net loss, despite the prevailing major downtrend.

On balance, in the market example illustrated in Figures 17.1 through 17.5, the benefit of filtering out some losing trades far outweighs the cumulative negative impact of the worse entries that result from using nonzero values for *N*1 and *N*2: For the entire period, the second parameter set generates a cumulative profit of $0.488 per pound ($12,200 per contract) versus a cumulative loss of –$0.379 per pound (–$9,475 per contract) for the first parameter set.

Although in some cases parameter sets with more sensitive entry conditions will experience the better performance, the outcome in our example is more typical. Generally speaking, the parameter sets with more restrictive entry conditions will do better, as the benefit of reducing whipsaw trades outweighs the disadvantage of worse entries. Ironically, human nature will lead most traders, especially novices, to choose more sensitive parameter sets because they will be attracted by the better entries and smaller surrender of open profits on individual trades offered by these sets, failing to fully appreciate the cumulative impact of reduced bad trades—a trait characteristic of more restrictive parameter sets.

It should be emphasized the selected example was intended to illustrate the mechanics of the wide-ranging day system across varied market conditions, not to put the system in the best light. Therefore, this example deliberately contained both intervals of strong wins as well as whipsaw losses. Note that I could easily have made the system look much more impressive by selecting a market and time period with much smoother trends. Such cherry-picked illustrations are all too common in trading books, articles, web sites, and—especially—advertisements. We return to this subject in the discussion of "the well-chosen example" in Chapter 19.

Run-Day Breakout System

Basic Concept

Up and *down run days* were defined in Chapter 9. As was explained, run days tend to occur in strongly trending markets. In this system, buy reversal signals are generated when the market closes above the maximum true high of a specified number of prior down run days. Similarly, sell reversal signals are generated when the market closes below the minimum true low of a specified number of prior up run days. The idea is that the ability of the market to close opposite the extreme point defined by one or more such strongly trending days implies a trend reversal has occurred.

Trading Signals

Buy case. Reverse to long whenever *both* of the following two conditions are met:

1. The close is above the maximum true high of the most recent $N2$ down run days. (*Note:* Only the run day true highs are considered, not the true highs on the interim days.)
2. The most recent run day is an *up* run day. (Without this second condition, in some cases, the first condition in the sell case would result in an automatic reversal back to a short position.)

Sell case. Reverse to short whenever *both* of the following two conditions are met:

1. The close is below the minimum true low of the most recent $N2$ up run days. (*Note*: Only the run day true lows are considered, not the true lows on the interim days.)
2. The most recent run day is a *down* run day. (Without this second condition, in some cases, the first condition in the buy case would result in an automatic reversal back to a long position.)

Daily Checklist

To generate trading signals, perform the following three steps each day:

1. Check whether the trading day $N1$ days prior to the current day can be defined as an up or a down run day.[2] (Recall that a run day cannot be defined until the close $N1$ days after the run day.) Keep track of all run days and their true highs and true lows.
2. If short, check whether today's close is above the maximum true high of the past $N2$ down run days. If it is, check whether the most recent run day was an up run day. If it was, reverse from short to long.
3. If long, check whether today's close is below the minimum true low of the past $N2$ up run days. If it is, check whether the most recent run day was a down run day. If it was, reverse from long to short.

Parameters

***N*1.** The parameter used to define run days. For example, if $N = 3$, a day would be defined as an up run day if its true high was greater than the maximum true high of the prior three days and its true low was less than the minimum true low of the following three days.

***N*2.** The number of prior down run days used to compute the maximum true high that must be exceeded by a close for a buy signal. (Also, the number of prior up run days used to compute the minimum true low that must be penetrated by a close for a sell signal.)

---

[2] Although uncommon, a day can be *both* an up run day and down run day. This unusual situation will occur if a day's true high is greater than the true highs during the prior and subsequent $N1$ days, and its true low is lower than the true lows during the prior and subsequent $N1$ days. Days that fulfill both the up and down run day definitions are not considered run days.

TABLE 17.2 Parameter Set List

	N1	N2
1.	3	2
2.	3	3
3.	3	4
4.	3	5
5.	5	2
6.	5	3
7.	5	4
8.	7	2
9.	7	3
10.	7	4

Parameter Set List

Table 17.2 provides a sample parameter set list. Readers can use this list as is or adjust it as desired.

An Illustrated Example

To illustrate the mechanics of the run-day breakout system, Figures 17.6 through 17.9 show the buy and sell signals generated by the system for the parameter set $N1 = 5$ and $N2 = 4$ in the WTI crude oil market. Down run days are denoted by downward-pointing arrows and up run days by upward-pointing arrows.

A close below the minimum true low of the four most recent up run days triggers a sell signal in January 2014 (Figure 17.6). Note the second condition for a sell signal—that is, the most recent run day is a down run day—was fulfilled on the day of the signal. Had the signal occurred one day earlier, no trade would have been taken because the December 31, 2013, down run day (the first in the string of four) would not yet have been confirmed, and the most recent run day would have been the December 19 up run day. (Remember that each run day marked with an arrow can be confirmed only after five days have passed.)

A buy signal occurs in February 2014 when the market closes above the true high of the December 31 down run day (the maximum true high of that string of four down run days). The second condition is also met as the most recent run day is an up run day.

The market drifts higher into June (note the predominance of up run days versus down run days) and then begins to sag into July (Figure 17.7). The system next goes short in July when the market closes below the minimum true low of the prior four up run days. The system stays short through the entire ensuing downtrend, which is characterized by a tremendous predominance of down run days, eventually reversing nearly nine months later (in April 2015) on the close above the true high of the cluster of down run days in March (Figure 17.8). The system holds this position through June as the market moves sideways to slightly higher. The downturn in July generates a flurry of down run days, and the system turns short on July 22 with a close below the March 25 low (Figure 17.9). Note that

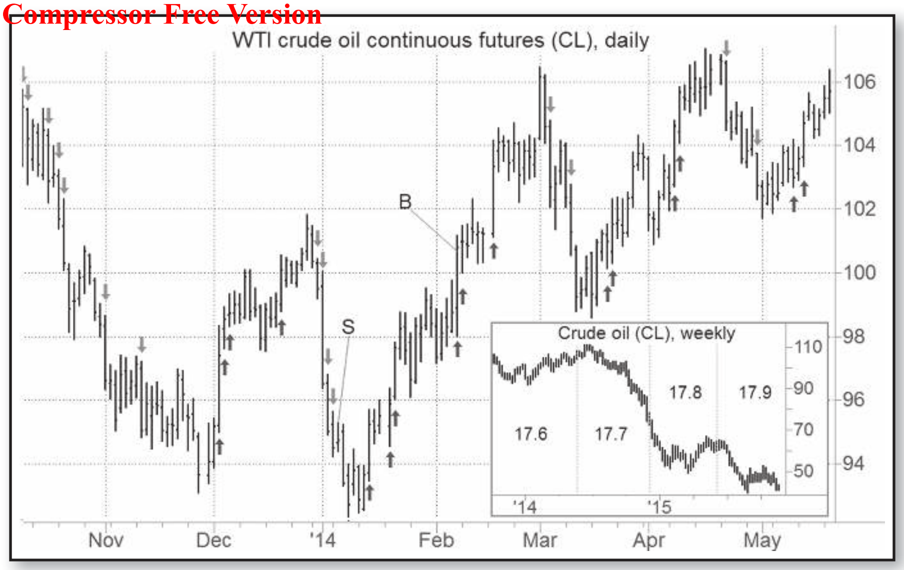

FIGURE 17.6 Run-Day Breakout System ($N1 = 5$; $N2 = 4$), Chart 1: WTI Crude Oil Continuous Futures

Note: The direction of the arrows indicates the direction of the run day.

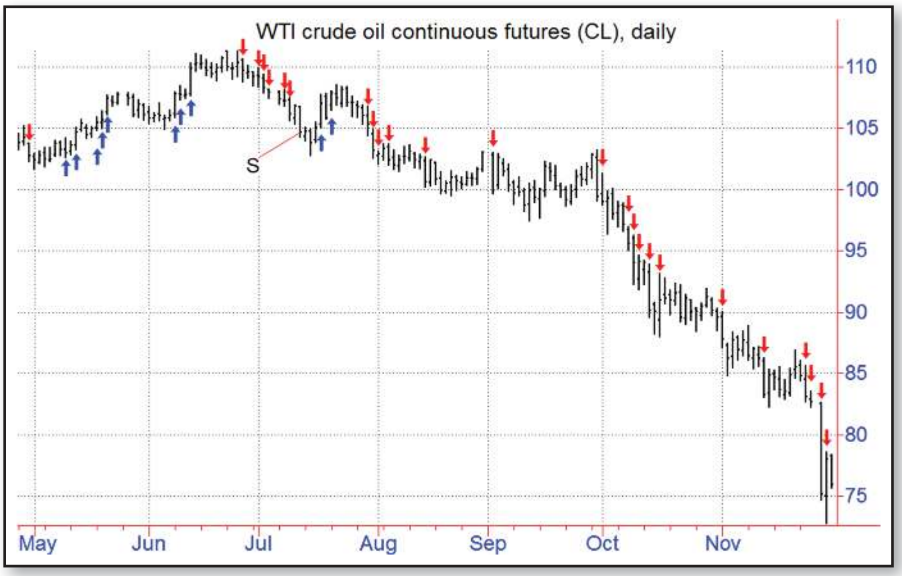

FIGURE 17.7 Run-Day Breakout System ($N1 = 5$; $N2 = 4$), Chart 2: WTI Crude Oil Continuous Futures

FIGURE 17.8 Run-Day Breakout System ($N1 = 5$; $N2 = 4$), Chart 3: WTI Crude Oil Continuous Futures

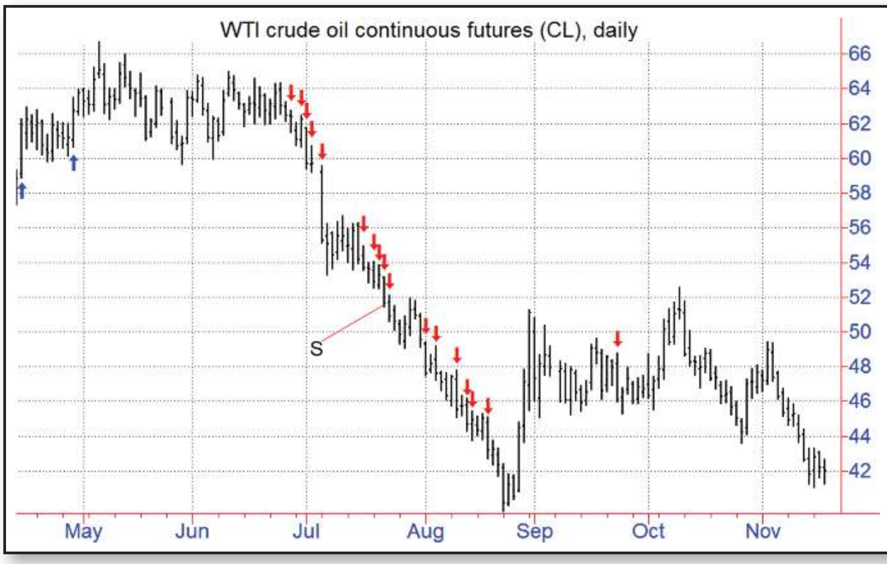

FIGURE 17.9 Run-Day Breakout System ($N1 = 5$, $N2 = 4$), Chart 4: WTI Crude Oil Continuous Futures

although the sharp rebound in late August is large enough to rally the market above the maximum true high of the four most recent down run days, there is no buy signal because there is no intervening up run day.

Overall, the system successfully exploits the major downtrend (July 2014 to August 2015) that occurs during the two-year survey period, capturing about half of the total profit that would be realized by a hypothetical trader who goes short at the high of the two-year period and covers the position at the low of the period. Readers, however, are cautioned against generalizing the system's performance based on this single market/single parameter set example. In most cases, the system will not attain the level of performance exhibited in this illustration.

Run-Day Consecutive Count System

Basic Concept

This system also uses run days as the key input in generating trading signals. In this system, reversal signals occur whenever there are a specified number of up run days without any intervening down run days, or vice versa.

Definitions

The system uses the following definitions:

Buy count. The buy count is activated whenever a sell signal is received. The count starts at zero and increases by one whenever a new up run day is defined. The count is reset to zero whenever there is a down run day. In effect, the buy count represents the number of up run days that occur without any intervening down run days. The buy count is closed when a buy signal is received.

Sell count. The sell count is activated whenever a buy signal is received. The count starts at zero and increases by one whenever a new down run day is defined. The count is reset to zero whenever there is an up run day. In effect, the sell count represents the number of down run days that occur without any intervening up run days. The sell count is closed when a sell signal is received.

Trading Signals

Buy case. Reverse to long whenever the buy count reaches $N2$. Keep in mind that the fulfillment of this condition will not be known until $N1$ days after the $N2$th *consecutive* up run day. (Consecutive here means that there are no intervening down run days, not that the up run days occur on consecutive days.)

Sell case. Reverse to short whenever the sell count reaches $N2$. Keep in mind that the fulfillment of this condition will not be known until $N1$ days after the $N2$th *consecutive* down run day. (Consecutive here means that there are no intervening up run days, not that the down run days occur on consecutive days.)

Daily Checklist

To generate trading signals, perform the following three steps each day:

1. Check whether the trading day $N1$ days prior to the current day can be defined as an up or a down run day. (Recall that a run day cannot be defined until the close $N1$ days after the run day.) If the day is an up run day, increase the buy count by one if the buy count is active (i.e., if the current position is short); otherwise, reset the sell count to zero. (Either the buy or sell count is always active, depending on whether the current position is short or long.) If the day is defined as a down run day, increase the sell count by one if the sell count is active (i.e., if current position is long); otherwise, reset the buy count to zero.
2. If the buy count is active, check whether it is equal to $N2$ after step 1. If it is, cover short, go long, close buy count, and activate sell count.
3. If the sell count is active, check whether it is equal to $N2$ after step 1. If it is, cover long, go short, close sell count, and activate buy count.

Parameters

***N*1.** The parameter used to define run days.
***N*2.** The number of *consecutive* run days required for a signal.

Parameter Set List

Table 17.3 provides a sample parameter set list. Readers can use this list as is or adjust it as desired.

TABLE 17.3 Parameter Set List

	*N*1	*N*2
1.	3	1
2.	3	2
3.	3	3
4.	3	4
5.	5	1
6.	5	2
7.	5	3
8.	7	1
9.	7	2
10.	7	3

An Illustrated Example

Figures 17.10 through 17.14 illustrate the signals generated by the run day consecutive count system for $N1 = 5$ and $N2 = 3$. In other words, the system reverses from long to short whenever there are three consecutive down run days and from short to long whenever there are three consecutive up run days. (Consecutive here means that there are no intervening run days in the opposite direction; not consecutive days.) *Keep in mind that the actual trade signal will not be received until the fifth close after the third consecutive run day, since a run day is not defined until* N*1 days after its occurrence (*N*1 = 5 in this example).*

The first signal in Figure 17.10—a buy in December 2013—occurs during a brief trading range and is reversed by a sell signal that occurs near the January 2014 low—a good example of how even profitable systems can generate terrible individual trade signals. The three consecutive up run days that start off February trigger a long position on February 12 (five days after the third up run day). Figure 17.11 shows this position remains intact until June 12, when the system reverses to short. Note that although the signal occurs on the fifth consecutive down run day in the sequence, it is the fact that this day is five days after the third consecutive down run day that triggers the trade. The downtrend that begins in June witnesses 18 down run days with no intervening up run days. In contrast, the preceding February–May uptrend contained 12 up run days and only one down run day.

The system reverses to the upside at the start of November 2014 (Figure 17.12), one week into a two-and-a-half-month trading range. The three consecutive down run days, which lead to the downside breakout of this range, turn the system short in January 2015. The next signal is the worst trade in the survey period, as the system reverses to long in February, shortly before the March 2 relative

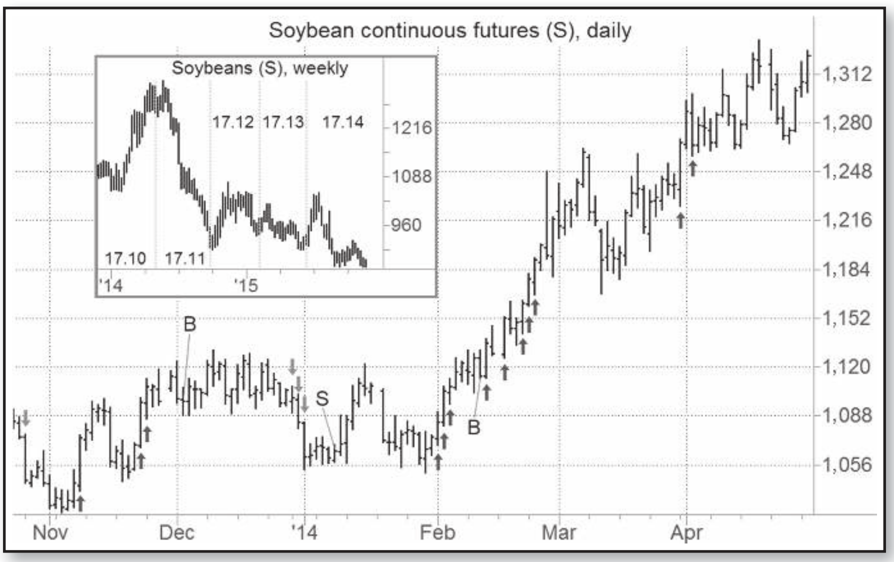

FIGURE 17.10 Run-Day Consecutive Count System, Chart 1: Soybean Continuous Futures

FIGURE 17.11 Run-Day Consecutive Count System, Chart 2: Soybean Continuous Futures

FIGURE 17.12 Run-Day Consecutive Count System, Chart 3: Soybean Continuous Futures

FIGURE 17.13 Run-Day Consecutive Count System, Chart 4: Soybean Continuous Futures

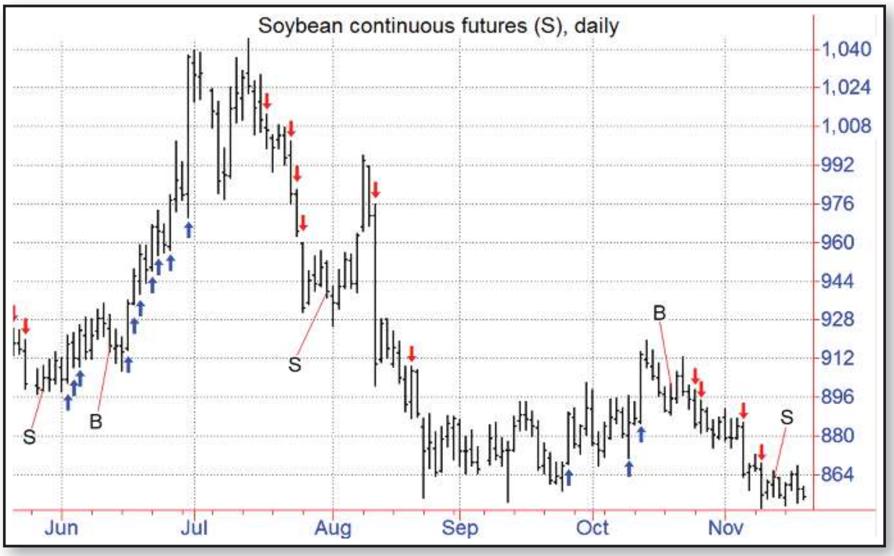

FIGURE 17.14 Run-Day Consecutive Count System, Chart 5: Soybean Continuous Futures

high (Figure 17.13). The system does not generate a sell signal until late May, just before a relative low. Fortunately, the system reverses back to long two weeks later in June 2015 just before a sharp, but short-lived, rally (Figure 17.14). The subsequent downside reversal is equally abrupt, and the system surrenders most of its profit on the long position by the time the next sell signal is generated in July. The final two signals occur in October and November 2015 within a relatively narrow consolidation phase.

It should be noted that our intention was to select a realistic market illustration of the system and not to cherrypick an example in which the system performed particularly well, as is typical in most books on trading. The foregoing example provided a market with both favorable (two 4-month trends) and unfavorable (a more than yearlong wide-swing trading range) price environments. On balance, the system was net profitable (a cumulative gain of 76.25 cents per bushel, or $3,812.50 per contract) as the profits during the two trending periods outweighed the losses during the extended trading range period.

Conclusion

In this chapter we have introduced some original trading systems. Although they are viable as described, readers may wish to experiment with modifications that use the concepts of these systems as the core of more complex approaches. The ultimate goal of this chapter was not to present specific trading systems, but rather to illustrate how basic chart concepts can be transformed into trading systems. The number of possible systems that can be constructed from the technical patterns and concepts already discussed in this volume are limited only by the imagination of the reader.

CHAPTER 18

Selecting the Best Futures Price Series for System Testing

Garbage in, garbage out.

—Anonymous

System traders wishing to test their ideas on futures prices have always faced a major obstacle: the transitory life span of futures contracts. In contrast to the equities market, where a given stock is represented by a single price series spanning the entire test period, in futures each market is represented by a string of expiring contracts. Proposed solutions to this problem have been the subject of many articles and a great deal of discussion. In the process, substantial confusion has been generated, as evidenced by the use of identical terms to describe different types of price series. Even worse, so much misinformation has been provided on this subject that many market participants now believe the equivalent of "the earth is flat" theory.

There are four basic types of price series that can be used. The definition, advantages, and disadvantages of each are discussed in turn.

Actual Contract Series

At a surface glance, the best route might seem to be simply to use the actual contract series. However, there are two major problems with this approach. First, if you are testing a system over a meaningful length of time, each market simulation will require a large number of individual price series. For example, a 15-year test run for a typical market would require using approximately 60 to 90 individual contract price series. Moreover, using the individual contract series requires an algorithm for determining what action to take at the rollover points. As an example of the type of problem that may be encountered, it is entirely possible for a given system to be long in the old contract and short in the new contract or vice versa. These problems are hardly insurmountable, but they make the use of individual contract series a somewhat unwieldy approach.

The awkwardness involved in using a multitude of individual contracts is not, however, the main problem. The primary drawback in using individual contract series is that the period of meaningful liquidity in most contracts is very short—much shorter than the already limited contract life spans. To see the scope of this problem, examine a cross section of futures price charts depicting the price action in the one-year period prior to expiration. In many markets, contracts don't achieve meaningful liquidity until the final five or six months of trading, and sometimes even less. This problem was illustrated in Chapter 5. The limited time span of liquid trading in individual contracts means that any technical system or method that requires looking back at more than about six months of data—as would be true for a whole spectrum of longer-term approaches—cannot be applied to individual contract series. Thus, with the exception of short-term system traders, the use of individual contract series is not a viable alternative. It's not merely a matter of the approach being difficult but, rather, its being impossible because the necessary data simply do not exist.

Nearest Futures

The problems in using individual contract series as just described has led to the construction of various linked price series. The most common approach is almost universally known as *nearest futures.* This price series is constructed by taking each individual contract series until its expiration and then continuing with the next contract until its expiration, and so on. This approach may be useful for constructing long-term price charts for purposes of chart analysis, but it is worthless for providing a series that can be used in the computer testing of trading systems.

The problem in using a nearest futures series is that there are price gaps between expiring and new contracts—and quite frequently these gaps can be very substantial. For example, assume the July corn contract expires at $4 and that the next nearest contract (September) closes at $3.50 on the same day. Assume that on the next day September corn moves from $3.50 to $3.62. A nearest futures price series will show the following closing levels on these two successive days: $4, $3.62. In other words, the nearest futures contract would imply a 38-cent loss on a day on which longs would have enjoyed (or shorts would have suffered) a price gain of 12 cents. This example is by no means artificial. In fact, it would be easy to find a plethora of similarly extreme situations in actual price histories. Moreover, even if the typical distortion at rollover is considerably less extreme, the point is that there is virtually always some distortion, and the cumulative effect of these errors would destroy the validity of any computer test.

Fortunately, few traders are naive enough to use the nearest futures type of price series for computer testing. The two alternative linked price series described in the next sections have become the approaches employed by most traders wishing to use a single price series for each market in computer testing.

Constant-Forward ("Perpetual") Series

The constant-forward (also known as "perpetual") price series consists of quotes for prices a constant amount of time forward. The interbank currency market offers actual examples of constant-forward price series. For example, the three-month forward price series for the euro represents the quote for the euro three months forward from each given day in the series. This is in contrast to the standard U.S. futures contract, which specifies a fixed expiration date.

A constant-forward series can be constructed from futures price data through interpolation. For example, if we were calculating a 90-day constant-forward (or perpetual) series and the 90-day forward date fell exactly one-third of the way between the expirations of the nearest two contracts, the constant-forward price would be calculated as the sum of two-thirds of the nearest contract price and one-third of the subsequent contract price. As we moved forward in time, the nearer contract would be weighted less, and the weighting of the subsequent contract would increase proportionately. Eventually, the nearest contract would expire and drop out of the calculation, and the constant-forward price would be based on an interpolation between the subsequent two contracts.

As a more detailed example, assume you want to generate a 100-day forward price series based on euro futures, which are traded in March, June, September, and December contracts. To illustrate the method for deriving the 100-day constant-forward price, assume the current date is January 20. In this case, the date 100 days forward is April 30. This date falls between the March and June contracts. Assume the last trading dates for these two contracts are March 14 and June 13, respectively. Thus, April 30 is 47 days after the last trading day for the March contract and 44 days before the last trading day for the June contract. To calculate the 100-day forward price for January 20, an average price would be calculated using the quotes for March and June euro futures on January 20, weighting each quote in inverse proportion to its distance from the 100-day forward date (April 30). Thus, if on January 20 the closing price of March futures is 130.04 and the closing price of June futures is 130.77, the closing price for the 100-day forward series would be:

$$\frac{44}{91}(130.04) + \frac{47}{91}(130.77) = 130.42$$

Note that the general formula for the weighting factor used for each contract price is:

$$W_1 = \frac{C_2 - F}{C_2 - C_1} \qquad W_2 = \frac{F - C_1}{C_2 - C_1}$$

where C_1 = number of days until the nearby contract expiration
C_2 = number of days until the forward contract expiration
F = number of days until forward quote date
W_1 = weighting for nearby contract price quote
W_2 = weighting for forward contract price quote

So, for example, the weightings of the March and June quotes that would be used to derive a 100-day forward quote on March 2 would be as follows:

$$\text{Weighting for March quote} = \frac{103 - 100}{103 - 12} = \frac{3}{91}$$

$$\text{Weighting for June quote} = \frac{100 - 12}{103 - 12} = \frac{88}{91}$$

As we move forward in time, the nearer contract is weighted less and less, but the weighting for the subsequent contract increases proportionately. When the number of days remaining until the expiration of the forward contract equals the constant-forward time (100 days in this example), the quote for the constant-forward series would simply be equal to the quote for the forward contract (June). Subsequent price quotes would then be based on a weighted average of the June and September prices. In this manner, one continuous price series could be derived.

The constant-forward price series eliminates the problem of huge price gaps at rollover points and is certainly a significant improvement over a nearest futures price series. However, this type of series still has major drawbacks. To begin, it must be stressed that one cannot literally trade a constant-forward series, since the series does not correspond to any real contract. An even more serious deficiency of the constant-forward series is that it fails to reflect the effect of the evaporation of time that exists in actual futures contracts. This deficiency can lead to major distortions—particularly in carrying-charge markets.

To illustrate this point, consider a hypothetical situation in which spot gold prices remain stable at approximately \$1,200/ounce for a one-year period, while forward futures maintain a constant premium of 1 percent per two-month spread. Given these assumptions, futures would experience a steady downtrend, declining \$73.82/ounce[1] (\$7,382 per contract) over the one-year period (the equivalent of the cumulative carrying-charge premiums). Note, however, the constant-forward series would completely fail to reflect this bear trend because it would register an approximate constant price. For example, a two-month constant-forward series would remain stable at approximately \$1,212/ounce ($1.01 \times \$1,200 = \$1,212$). Thus, the price pattern of a constant-forward series can easily deviate substantially from the pattern exhibited by the actual traded contracts—a highly undesirable feature.

Continuous (Spread-Adjusted) Price Series

The spread-adjusted futures series, commonly known as *continuous futures,* is constructed to eliminate the distortions caused by the price gaps between consecutive futures contracts at their transition points. In effect, the continuous futures price will precisely reflect the fluctuations of a futures position that is continuously rolled over to the subsequent contract N days before the last trading day, where N is a parameter that needs to be defined. If constructing their own continuous futures data series, traders should select a value of N that corresponds to their actual trading practices.

[1] This is true since, given the assumptions, the one-year forward futures price would be approximately \$1,273.82 ($1.01^6 \times \$1,200 = \$1,273.82$) and would decline to the spot price (\$1,200) by expiration.

For example, if a trader normally rolls a position over to a new contract approximately 20 days before the last trading day, *N* would be defined as 20. The scale of the continuous futures series is adjusted so the current price corresponds to a currently traded futures contract.

Table 18.1 illustrates the construction of a continuous futures price for the soybean market. For simplicity, this example uses only two contract months, July and November; however, a continuous price could be formed using any number of traded contract months. For example, the continuous futures price could be constructed using the January, March, May, July, August, September, and November soybean contracts.

TABLE 18.1 **Construction of a Continuous Futures Price Using July and November Soybeans (cents/bushel)***

Date	Contract	Actual Price	Spread at Rollover (Nearby Forward)	Cumulative Adjustment Factor	Unadjusted Continuous Futures (Col. 3 + Col. 5)	Continuous Futures Price (Col. 6 – 772.5)
6/27/12	Jul-12	1,471			1,471	698.5
6/28/12	Jul-12	1,466			1,466	693.5
6/29/12	Jul-12	1,512.75			1,512.75	740.25
7/2/12	Nov-12	1,438	85	85	1,523	750.5
7/3/12	Nov-12	1,474.75		85	1,559.75	787.25
* * *						
10/30/12	Nov-12	1,533.75		85	1,618.75	846.25
10/31/12	Nov-12	1,547		85	1,632	859.5
11/1/12	Jul-13	1,474	86.25	171.25	1,645.25	872.75
11/2/12	Jul-13	1,454		171.25	1,625.25	852.75
* * *						
6/27/13	Jul-13	1,548.5		171.25	1,719.75	947.25
6/28/13	Jul-13	1,564.5		171.25	1,735.75	963.25
7/1/13	Nov-13	1,243.25	312.5	483.75	1,727	954.5
7/2/13	Nov-13	1,242.5		483.75	1,726.25	953.75
* * *						
10/30/13	Nov-13	1,287.5		483.75	1,771.25	998.75
10/31/13	Nov-13	1,280.25		483.75	1,764	991.5
11/1/13	Jul-14	1,224.5	45.5	529.25	1,753.75	981.25
11/4/13	Jul-14	1,227.75		529.25	1,757	984.5
* * *						
6/27/14	Jul-14	1,432		529.25	1,961.25	1,188.75
6/30/14	Jul-14	1,400.5		529.25	1,929.75	1,157.25
7/1/14	Nov-14	1,147.5	243.25	772.5	1,920	1,147.5
7/2/14	Nov-14	1,141.5		772.5	1,914	1,141.5

*Assumes rollover on last day of the month preceding the contract month.

For the moment, ignore the last column in Table 18.1 and focus instead on the unadjusted continuous futures price (column 6). At the start of the period, the actual price and the unadjusted continuous futures price are identical. At the first rollover point, the forward contract (November 2012) is trading at an 85-cent discount to the nearby contract (July 2012). All subsequent prices of the November 2012 contract are then adjusted upward by this amount (the addition of a positive nearby/forward spread), yielding the unadjusted continuous futures prices shown in column 6. At the next rollover point, the forward contract (July 2013) is trading at an 86.25-cent discount to the nearby contract (November 2012). As a result, all subsequent actual prices of the July 2013 contract must now be adjusted by the cumulative adjustment factor—the total of all rollover gaps up to that point (171.25 cents)—in order to avoid any artificial price gaps at the rollover point. This cumulative adjustment factor is indicated in column 5. The unadjusted continuous futures price is obtained by adding the cumulative adjustment factor to the actual price.

The preceding process is continued until the current date is reached. At this point, the final cumulative adjustment factor is subtracted from all the unadjusted continuous futures prices (column 6), a step that sets the current price of the series equal to the price of the current contract (November 2014 in our example) without changing the shape of the series. This continuous futures price is indicated in column 7 of Table 18.1. Note that although actual prices seem to imply a net price decline of 329.50 cents during the surveyed period, the continuous futures price indicates a 443-cent *increase*—the actual price change that would have been realized by a constant long futures position.

In effect, the construction of the continuous series can be thought of as the mathematical equivalent of taking a nearest futures chart, cutting out each individual contract series contained in the chart, and pasting the ends together (assuming a continuous series employing all contracts and using the same rollover dates as the nearest futures chart).

In some markets, the spreads between nearby and forward contracts will range from premiums to discounts (e.g., cattle). However, in other markets, the spread differences will be unidirectional. For example, in the gold market, the forward month always trades at a premium to the nearby month.[2] In these types of markets, the spread-adjusted continuous price series can become increasingly disparate from actual prices.

It should be noted that when nearby premiums at contract rollovers tend to swamp nearby discounts, it is entirely possible for the series to eventually include negative prices for some past periods as cumulative adjustments mount, as illustrated in the soybean continuous futures chart in Figure 18.1. The price gain that would have been realized by a continuously held futures position during this period

[2] The reason for this behavioral pattern in gold spreads is related to the fact that world gold inventories exceed annual usage by many multiples, perhaps even by as much as a hundredfold. Consequently, there can never actually be a "shortage" of gold—and a shortage of nearby supplies is the only reason why a storable commodity would reflect a premium for the nearby contract. (Typically, for storable commodities, the fact that the forward contracts embed carrying costs will result in these contracts trading at a premium to more nearby months.) Gold prices fluctuate in response to shifting perceptions of gold's value among buyers and sellers. Even when gold prices are at extremely lofty levels, it does not imply any actual shortage, but rather an upward shift in the market's perception of gold's value. Supplies of virtually any level are still available—at some price. This is not true for most commodities, in which there is a definite relevant limit in total supplies.

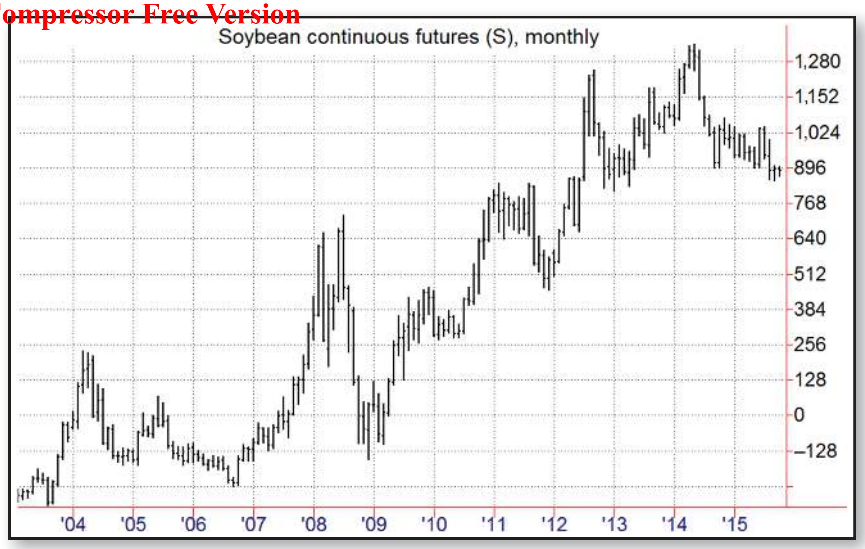

FIGURE 18.1 "Negative" Prices in a Continuous Futures Chart: Soybean Continuous Futures
Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

far exceeded the net price gain implied by nearest futures, and the subtraction of the cumulative adjustment factor from the most recent (2015) prices would result in negative prices for the majority of the time before 2009. Such an outcome is unavoidable if the continuous futures price series is to reflect the net gain in a continually held long position and if the series is shifted by the constant factor necessary to set the current continuous futures price equal to the current contract actual price.

Although the fact that a continuous futures price series could include negative prices may sound disconcerting, it does not present any problems in using the series for testing systems. The reason for this is that in measuring the profits or losses of trades, it is critical that the price series employed accurately reflects price *changes,* not price *levels.* However, it also will often be useful to generate the actual prices that correspond to the continuous futures prices in order to facilitate such applications as checking trading signals against actual contract charts.

It should also be noted that the transition between contracts need not occur on the last trading day, as is the conventional assumption in the nearest futures price series. In fact, because physically delivered contracts are particularly vulnerable to distortions in their final weeks of trading due to technical concerns regarding delivery, it probably makes sense to avoid these prices in constructing a continuous series. It follows, then, that one should use a rollover date before the last trading day (e.g., 20 days prior to the last trading day).

Comparing the Series

It is important to understand that a linked futures price series can only accurately reflect either price *levels,* as do nearest futures, or price *moves*, as do continuous futures, but not both—much as a coin can land on either heads or tails but not both. The adjustment process used to construct continuous series means that past prices in a continuous series will not match the actual historical prices that prevailed at the time. However, the essential point is that the continuous series is the only linked futures series that will exactly reflect price swings and hence equity fluctuations in an actual trading account. Consequently, it is the only linked series that can be used to generate accurate simulations in computer testing of trading systems.

The preceding point is absolutely critical! Mathematics is not a matter of opinion. There is one right answer and there are many wrong answers. The simple fact is that if a continuous futures price series is defined so that rollovers occur on days consistent with rollovers in actual trading, results implied by using this series will precisely match results in actual trading (assuming, of course, accurate commission and slippage cost estimates). In other words, the continuous series will exactly parallel the fluctuations of a constantly held (i.e., rolled over) long position. All other types of linked series will not match actual market price movements.

To illustrate this statement, we compare the implications of various price series using the sideways gold market example cited earlier in this chapter (i.e., gold hovering near $1,200 and a forward/nearby contract premium equal to 1 percent per two-month spread). A trader buying a one-year forward futures contract would therefore pay approximately $1,273.82 ($1.01^6 \times \$1,200 = \$1,273.82$). The spot price would reflect a sideways pattern near $1,200. As previously seen, a 60-day constant-forward price would reflect a sideways pattern near $1,212 ($1.01 \times \$1,200$). A nearest futures price series would exhibit a general sideways pattern, characterized by extended minor downtrends (reflecting the gradual evaporation of the carrying charge time premium as each nearby contract approached expiration), interspersed with upward gaps at rollovers between expiring and subsequent futures contracts.

Thus the spot, constant-forward, and nearest futures price series would all suggest that a long position would have resulted in a break-even trade for the year. In reality, however, the buyer of the futures contract pays $1,273.82 for a contract that eventually expires at $1,200. Thus, from a trading or real-world viewpoint, the market actually witnesses a downtrend. The continuous futures price is the only price series that reflects the market decline—and real dollar loss—a trader would actually have experienced.

I have often seen comments or articles by industry "experts" arguing for the use of constant-forward (perpetual) series instead of continuous series in order to avoid distortions. This argument has it exactly backwards. Whether these proponents of constant-forward series adopt their stance because of naïveté or self-interest (i.e., they are vendors of constant-forward-type data), they are simply wrong. This is not a matter of opinion. If you have any doubts, try matching up fluctuations in an actual trading account with those that would be implied by constant-forward-type price series. You will soon be a believer.

Are there any drawbacks to the continuous futures time series? Of course. It may be the best solution to the linked series problem, but it is not a perfect answer. A perfect alternative simply does not exist. One potential drawback, which is a consequence of the fact that continuous futures accurately reflect only price swings, not price levels, is that continuous futures cannot be used for any type of percentage calculations. This situation, however, can be easily remedied. If a system requires the calculation of a percentage change figure, use continuous futures to calculate the nominal price change and nearest futures for the divisor. Also, there is some unavoidable arbitrariness involved in constructing a continuous series, since one must decide which contracts to use and on what dates the rollovers should occur. However, this issue is not really a problem since these choices should merely mirror the contracts and rollover dates used in actual trading. Moreover, there is arbitrariness involved in the use of any of the price series discussed. Finally, in some markets, the contracts being linked together may have very different past price patterns (as is often the case in livestock markets). However, this problem would exist in any kind of linked series.

Conclusion

For the purpose of computer testing of trading systems, there are only two types of valid price series: (1) individual contract series and (2) continuous futures series. Individual contract series are a viable approach only if the methodologies employed do not require looking back more than four or five months in time (a restriction that rules out a vast number of technical approaches). In addition, the use of individual contract series is far clumsier. Thus, for most purposes, the continuous futures price series provides the best alternative. As long as one avoids using continuous prices for percentage calculations, this type of price series will yield accurate results (i.e., results that parallel actual trading) as well as provide the efficiency of a single series per market. Again, I would strongly caution data users to avoid being misled by those who argue for the use of constant-forward-type series in computer testing applications. If your goal is a price series that will accurately reflect futures trading, the constant-forward series will create distortions rather than avoid them.

Over the years more—but not all—data vendors and system-testing platforms have embraced the continuous futures series described here as the default data type for long-term analysis and system testing. Traders should nonetheless confirm with the vendor that long-term futures data series linking different contracts are indeed constructed using the continuous futures (i.e., spread-adjusted) methodology. Traders should also be cognizant of the contracts and rollover dates used by the vendor so that they can match their contract selection and rollover dates accordingly. Vendors should be able to provide a clear explanation of the methodology they employ for constructing long-term (i.e., linked-contract) futures data series.

CHAPTER 19

Testing and Optimizing Trading Systems

Every decade has its characteristic folly, but the basic cause is the same: people persist in believing that what has happened in the recent past will go on happening into the indefinite future, even while the ground is shifting under their feet.

—George J. Church

The Well-Chosen Example[1]

You've plunked down your $895 to attend the 10th annual "Secret of the Millionaires" futures trading seminar. At that price, you figure the speakers will be revealing some very valuable information.

The current speaker is explaining the Super-Razzle-Dazzle (SRD) commodity trading system. The slide on the huge screen reveals a price chart with "B" and "S" symbols representing buy and sell points. The slide is impressive: All of the buys seem to be lower than the sells.

This point is brought home even more dramatically in the next slide, which reveals the equity stream that would have been realized trading this system—a near-perfect uptrend. Not only that but the system is also very easy to keep up.

As the speaker says, "All it takes is 10 minutes a day and a knowledge of simple arithmetic."

You never realized making money in futures could be so simple. You could kick yourself for not having attended the first through ninth annual seminars.

[1] The following section is adapted from an article that first appeared in *Futures* magazine in September 1984.

Once you get home, you select 10 diversified markets and begin trading the SRD system. Each day you plot your equity. As the months go by, you notice a strange development. Although the equity in your account exhibits a very steady trend, just as the seminar example did, there is one small difference: The trend on your equity chart is down. What went wrong?

The fact is you can find a favorable illustration for almost any trading system. The mistake is in extrapolating probable future performance on the basis of an isolated and well-chosen example from the past.

A true-life example may help illustrate this point. Back in 1983, when I had been working on trading systems for only a couple of years, I read an article in a trade magazine that presented the following very simple trading system:

1. If the six-day moving average is higher than the previous day's corresponding value, cover short and go long.
2. If the six-day moving average is lower than the previous day's corresponding value, cover long and go short.

The article used the Swiss franc in 1980 as an illustration. Without going into the details, suffice it to say that applying this system to the Swiss franc in 1980 would have resulted in a profit of $17,235 per contract after transaction costs. Even allowing for a conservative fund allocation of $6,000 per contract, this implied an annual gain of 287 percent! Not bad for a system that can be summarized in two sentences. It is easy to see how traders, presented with such an example, might eagerly abandon their other trading approaches for this apparent money machine.

I couldn't believe such a simple system could do so well. So I decided to test the system over a broader period—1976 to mid-1983[2]—and a wide group of markets.

Beginning with the Swiss franc, I found that the total profit during this period was $20,473. In other words, excluding 1980, the system made only $3,238 during the remaining 6½ years. Thus, assuming that you allocated $6,000 to trade this approach, the average annual percent return for those years was a meager 8 percent—quite a comedown from 287 percent in 1980.

But wait. It gets worse. Much worse.

When I applied the system to a group of 25 markets from 1976 through mid-1983, the system lost money in 19 of the 25 markets. In 13 of the markets—more than half of the total survey—the loss exceeded $22,500, or $3,000 per year, per contract! In five markets, the loss exceeded $45,000, equivalent to $6,000 per year, per contract! Also, it should be noted that, even in the markets where the system was profitable, its performance was well below gains exhibited for these markets during the same period by most other trend-following systems.

There was no question about it. This was truly a bad system. Yet if you looked only at the well-chosen example, you might think you had stumbled upon the trading system Jesse Livermore used in his good years. Talk about a gap between perception and reality.

This system witnessed such large, broadly based losses that you may well wonder why fading the signals of such a system might not provide an attractive trading strategy. The reason is that most of the

[2]The start date was chosen to avoid the distortion of the extreme trends witnessed by many commodity markets during 1973–1975. The end date merely reflected the date on which I tested this particular system.

losses are the result of the system being so sensitive that it generates large transaction costs. (Transaction costs include commission costs *plus* slippage. The concept of slippage is discussed later in this chapter.) This sensitivity of the system occasionally is beneficial, as was the case for the Swiss franc in 1980. However, on balance, it is the system's major weakness.

Losses due to transaction costs would not be realized as gains by fading the system. Moreover, doing the opposite of all signals would generate equivalent transaction costs. Thus, once transaction costs are incorporated, the apparent attractiveness of a contrarian approach to using the system evaporates.

Because the related episode and the system testing it inspired occurred many years ago, some readers might justifiably wonder whether the system has been a viable strategy in more recent years. To answer this question, we tested the same system on a portfolio of 31 U.S. futures contracts for the 10 years ending November 30, 2015, and produced similar results: Only 12 of the 31 markets generated a net gross profit—that is, a profit before accounting for commissions or slippage. Incorporating a \$25 commission and slippage assessment reduced the number of profitable markets to nine, and the total losses of the unprofitable markets outweighed the profits of the winning markets by a factor of more than 4 to 1, with a total cumulative loss of −\$940,612 for the entire 10-year period (assuming a trade size of one contract per market).

The moral is simple: Don't draw any conclusions about a system (or indicator) on the basis of isolated examples. The only way you can determine if a system has any value is by testing it (without benefit of hindsight) over an extended time period for a broad range of markets.

Basic Concepts and Definitions

A *trading system* is a set of rules that can be used to generate trade signals. A *parameter* is a value that can be freely assigned in a trading system in order to vary the timing of signals. For example, in the basic breakout system, N (the number of prior days whose high or low must be exceeded to indicate a signal) is a parameter. Although the operation of the rules in the system will be identical whether $N = 7$ or $N = 40$, the timing of the signals will be vastly different. (For an example, see Figure 16.5 in Chapter 16.)

Most trading systems will have more than one parameter. For example, in the crossover moving average system there are two parameters: the length of the short-term moving average and the length of the long-term moving average. Any combination of parameter values is called a *parameter set*. For example, in a crossover moving average system, moving averages of 10 and 40 would represent a specific parameter set. Any other combination of moving average values would represent another parameter set. In systems with only one parameter (e.g., breakout), the parameter set would consist of only one element.[3]

[3] Note that the terms *parameter set* and *system variation* (the latter was used in Chapter 16) refer to identical concepts. The introduction of the term *parameter set* was merely deferred until this chapter because doing so allowed for a more logically ordered presentation of the material.

Most "generic" systems are limited to one or two parameters. However, the design of more creative and flexible systems, or the addition of modifications to basic systems, will usually imply the need for three or more parameters. For example, adding a confirmation time delay rule to the crossover moving average system would imply a third parameter: the number of days in the time delay.

As a general principle, it is wise to use the simplest form of a system (i.e., the least number of parameters) that does not imply any substantial deterioration in performance relative to the more complex versions. However, one should not drop parameters that are deemed important simply to reduce the number of implied parameter sets. In this case, a more reasonable approach would be to limit the number of parameter sets actually tested.

It should be noted that even for a simple one- or two-parameter-set system, it is not necessary to test all possible combinations. For example, in a simple breakout system in which one wishes to test the performance for values of $N = 1$ to $N = 100$, it is not necessary to test each integer in this range. A more efficient approach would be to first test the system using spaced values for N (e.g., 10, 20, 30, . . . , 100), and then, if desired, the trader could focus on any areas that appeared to be of particular interest. For example, if the system exhibited particularly favorable performance for the parameter values $N = 40$ and $N = 50$, the trader might want to also test some other values of N in this narrower range. Such an additional step, however, is probably unnecessary, since, as is discussed later in this chapter, performance differences in parameter set values—particularly values in such close proximity—are probably a matter of chance and lack any significance.

As a more practical real-life example, assume we wish to test a crossover moving average system that includes a time-delay confirmation rule. If we were interested in the performance of the system for parameter values 1 to 50 for the shorter-term moving average, 2 to 100 for the longer-term moving average, and 1 to 20 for the time delay, there would be a total of 74,500 parameter sets.[4] Note that we cannot reduce the number of parameters without severely damaging the basic structure of the system. However, we can test a far more limited number of parameter sets and still produce a very good approximation of the system's overall performance. Specifically, we might use increments of 10 for the shorter-term moving average (10, 20, 30, 40, and 50), increments of 20 for the longer-term moving average (20, 40, 60, 80, and 100), and three selected values for the time delay (e.g., 5, 10, and 20). This approach would limit the number of parameter sets to be tested to 57.[5] Once these parameter sets are tested, the results would be analyzed, and a moderate number of additional parameter sets might be tested as suggested by this evaluation. For example, if a time delay of 5—the smallest value tested—seemed to work best for most favorably performing parameter sets, it would also be reasonable to test smaller values for the time delay.

Conceptually, it might be useful to define four types of parameters:

Continuous parameter. A continuous parameter can assume any value within a given range. A percentage price penetration would be an example of a continuous parameter. Because a

---

[4] To avoid double counting, each "short-term" moving average can only be combined with a "long-term" moving average for a longer period. Thus, the total number of combinations is given by $(99 + 98 + 97 + \ldots + 50)(20) = 74{,}500$.

[5] $(5 + 4 + 4 + 3 + 3)(3) = 57$.

continuous parameter can assume an infinite number of values, it is necessary to specify some interval spacing in testing such a parameter. For example, a percent penetration parameter might be tested over a range of 0.05 percent to 0.50 percent, at intervals of 0.05 (i.e., 0.05, 0.10, . . . , 0.50). It is reasonable to expect performance results to change only moderately for an incremental change in the parameter value (assuming a sufficiently long test period).

Discrete parameter. A discrete parameter can assume only integer values. For example, the number of days in a breakout system is a discrete parameter. Although one can test a discrete parameter for every integer value within the specified range, such detail is often unnecessary, and wider spacing is frequently employed. As with continuous parameters, it is reasonable to expect performance results to change only moderately for a small change in the parameter value.

Code parameter. A code parameter is used to represent a definitional classification. Thus, there is no significance to the cardinal value of a code parameter. For example, assume we wish to test a simple breakout system using three different definitions of a breakout (buy case): *close* above previous *N*-day *high, high* above previous *N*-day *high,* and *close* above previous *N*-day *high close.* We could test each of these systems separately, but it might be more efficient to use a parameter to specify the intended definition. Thus, a parameter value of 0 would indicate the first definition, a value of 1 the second definition, and a value of 2 the third definition. Note that there are only three possible values for this parameter, and that there is no significance to incremental changes in parameter values.

Fixed or nonoptimized parameter. Normally, any type of parameter will be allowed to assume different values in testing a system. However, in systems with a large number of parameters, it may be necessary to fix some parameter values in order to avoid an excessive number of parameter sets. Such parameters are called *nonoptimized parameters.* For example, in a nonsensitive (slow) trend-following system, we might wish to include a backup stop rule to prevent catastrophic losses. By definition, in this situation, the stop rule would be activated on only a few occasions. Consequently, any parameters implicit in the stop rule could be fixed, since variation in these parameter values would not greatly affect the results.

Choosing the Price Series

The first step in testing a system in a given market is choosing the appropriate price series. The issues related to this selection were fully detailed in Chapter 18. Generally speaking, a continuous futures series is the preferred choice, although actual contract data could be used for short-term trading systems.

Choosing the Time Period

Generally speaking, the longer the test period, the more reliable the results. If the time period is too short, the test will not reflect the system's performance for a reasonable range of market situations. For example, a test of a trend-following system on the Canadian dollar market that used only the three

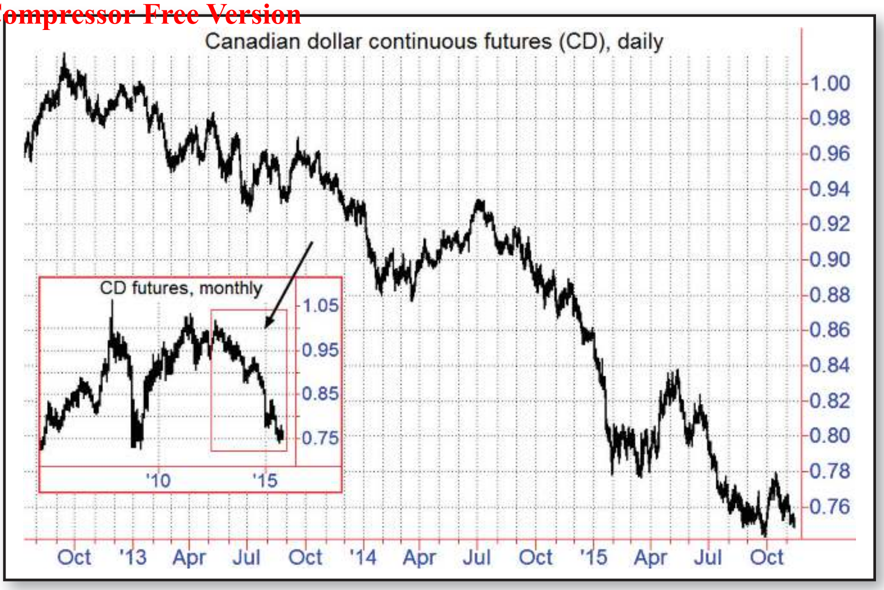

FIGURE 19.1 Major Trending Phase as Unrepresentative Price Sample: Canadian Dollar Continuous Futures

Chart created using TradeStation.

years of data from roughly October 2012 to October 2015—a period dominated by a sustained bear market (see Figure 19.1)—would yield highly misleading results in terms of the system's probable long-term performance, as evidenced by the monthly chart inset, which shows the market's price action dating back to 2004. Although testing over only the recent past is almost always undesirable, longer periods are not always necessarily better for testing than shorter ones. In some markets, if too long a period is used for testing a system, the earlier years in the survey period might be extremely unrepresentative of current market conditions.

Although it is impossible to provide a decisive answer as to the optimum number of years to be used in testing, 10 to 20 years is a reasonable range. For short-term trading systems (average duration of trades equal to a few weeks or less), a shorter test period (e.g., 5 to 10 years) would probably be sufficient. Trading system test results based on time periods significantly shorter than these guidelines should be suspect. In fact, it is rather incredible that some published studies on trading systems are based on test periods of two years or less.

Trading systems that use intraday data do not need to be tested over as long a time period as is the case for daily data because any time period will contain far more data points. For example, in the case of five-minute bars, a stock-index futures contract—just during the stock market's cash trading session—will generate the equivalent of a year's worth of daily price bars (252) in a little more than three days. A year's worth of this five-minute data would contain approximately as many price bars as 78 years of daily data.

However, the far greater amount of data inherent in intraday data does not mean the test period can be reduced proportionally—not even close. The governing principle will always be to select enough data to expose the system to a wide range of market conditions. A trader testing a system based on five-minute bars should run the test on far more than 30 days of data, even though this data contains more bars than 10 years of daily price bars, since the larger-scale market conditions can often be relatively static over such brief time periods. For example, the intraday price action during a very strong 30-day trending period will likely differ dramatically from the typical intraday price action during a 30-day trading range. The necessity that any meaningful system test span bull, bear, and sideways markets means that even intraday systems will need to be tested over a period of at least several years, if not more. In fact, given the current speed of computer processing, if the data are available, there is no compelling reason to run intraday systems tests for significantly shorter periods than daily systems. Sure, such tests will include dramatically more data, but that is a good thing.

Ideally, one should test a system using a longer time period (e.g., 15 years) and then evaluate the results for the period as a whole and various shorter time intervals (e.g., individual years). Such an approach is important in determining the system's degree of *time stability*—the relative performance consistency from one period to the next. Time stability is important because it enhances confidence regarding a system's potential for maintaining consistently favorable performance in the future. Most people would be quite hesitant about using a system that generated significant net profits over a 15-year period due to three spectacularly performing years but then witnessed losses or near break-even results in the remaining 12 years—and rightly so. In contrast, a system that registered moderate net gains during the 15-year period and was profitable in 14 of the 15 years would undoubtedly be viewed as more attractive by most traders.

Realistic Assumptions

System traders often discover that their actual results are substantially worse than the paper trading results implied by the system. In fact, this situation is so common that it even has its own name: *slippage*. Assuming that the divergence in the results is not due to errors in the program, slippage is basically a consequence of a failure to use realistic assumptions in testing the system. Basically, there are two types of such faulty assumptions:

1. **Transaction costs.** Most traders don't realize that merely adjusting for actual commission costs in testing a system is not a sufficiently rigid assumption. The reason for this is that commissions account for only a portion—and usually a minor portion—of transaction costs. Another less tangible, but no less real, cost is the difference between the theoretical execution price and the actual fill price. For example, if one is testing a system assuming order entry on the close, the use of the midpoint of the closing range might not be a realistic assumption. For some reason, buys near the upper end of the closing range and sells near the lower end of the closing range seem to be far more common than their reverse counterparts. There are two ways of addressing this problem. First, use the worst possible fill price (e.g., high of the closing range for buys). Second, use a transaction cost per trade assumption much greater than the actual

historical commission costs (e.g., $25 per side, per trade). The latter approach is preferable because it is more general. For example, how would one decide the worst possible fill price for an intraday stop order?

2. **Limit days.** Unless it is programmed otherwise, an automated trading system will indicate executions on the receipt of each signal. However, in the real world, things are not quite so simple. Occasionally, execution will not be possible because the market is locked at the daily permissible limit. Or even if execution is possible, it could occur at a much worse level than the intended price because the market gaps far beyond the signal trigger price. Although nearly continuous trading hours have made these events less common than in decades past, they still occur, especially in less liquid markets. If one assumes execution in such a situation, the paper results may dramatically overstate actual performance. Figure 19.2 illustrates the difference even a single locked-limit day can have on trade results. September 2011 corn futures closed limit down at 648 cents on June 30. A trader who wanted—or worse, needed—to sell on this close but did not receive a fill would have had to wait for the next session to execute the trade. The market opened 41.25 cents lower the next day, representing a $2,062.50 loss per contract, assuming the trade was filled exactly at the opening price.

The potential systems trader may discover that seemingly attractive trading systems disintegrate once realistic assumptions are employed. This characteristic is particularly true for very active systems, which generate very large transaction costs. However, it is far better to make this discovery in the analytical testing stage than in actual trading.

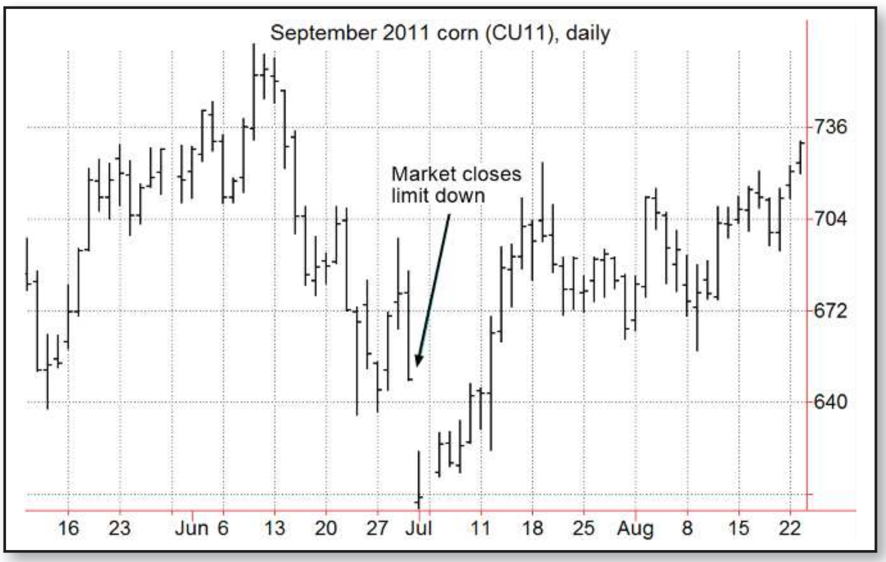

FIGURE 19.2 Wide Gap between Signal Price and Actual Entry: Impact of Limit Days (September 2011 Corn)

Optimizing Systems

Optimization refers to the process of finding the best-performing parameter set(s) for a given system applied to a specific market. The underlying premise of optimization is that parameter sets that worked best in the past have a greater probability of superior performance in the future. (The question of whether this assumption is valid is addressed in the next section.)

A basic question that must be considered in optimization is what criteria should be used for defining best performance. Frequently, best performance is simply interpreted as largest equity gain. However, such a definition is incomplete. Ideally, four factors should be considered in performance comparisons:

1. **Percent return.** Return measured relative to funds needed to trade the system. The importance of using percent return rather than nominal gain is detailed in Chapter 20.
2. **Risk measure.** In addition to percent gain, it is also important to employ some measure of equity fluctuations (e.g., variability in rate of gain, retracements in equity). Besides the obvious psychological reasons for wishing to avoid parameter sets and systems with high volatility, a risk measure is particularly significant because one might pick an unfavorable starting date for trading the system. Chapter 20 discusses several performance measures that incorporate both percent return and risk.
3. **Parameter stability.** It is not sufficient to find a parameter set that performs well. It is also necessary to ascertain the parameter set does not reflect a fluke in the system. In other words, we wish to determine that similar parameter sets also exhibit favorable performance. In fact, the goal of optimization should be to find broad regions of good performance rather than the single best-performing parameter set.

   For example, if in testing a simple breakout system one found that the parameter set $N = 7$ exhibited the best percent return/risk characteristics but that performance dropped off very sharply for parameter sets $N < 5$ and $N > 9$, while all sets in the range $N = 25$ to $N = 54$ performed relatively well, it would make much more sense to choose a parameter set from the latter range. Why? Because the exceptional performance of the set $N = 7$ appears to be a peculiarity of the historical price data, which is not likely to be repeated. The fact that surrounding parameter sets performed poorly suggests that there is no basis for confidence in trading the parameter set $N = 7$. In contrast, the broad range of performance stability for sets in the region $N = 25$ to $N = 54$ suggests that a set drawn from the center of this range would have a better prospect for success.
4. **Time stability.** As detailed in a previous section, it is important to ascertain that favorable performance for the period as a whole is truly representative of the total period rather than a reflection of a few isolated intervals of extraordinary performance.

For comparisons involving different parameter sets for the *same* system, the preceding factors tend to be highly correlated. Generally, the parameter sets with the best gains will also be the sets that exhibit the smallest equity retracements. Consequently, for the optimization of a single system, the use of a basic return/risk measure (e.g., the Sharpe ratio or the gain-to-pain ratio) will usually yield similar results to a complex performance evaluation that incorporates multiple performance measures. Thus, although the multifactor performance evaluation is theoretically preferable, it is often not essential. However, if one is comparing parameter sets from completely different systems, the explicit consideration of risk, parameter stability, and time stability is more important.

The foregoing represents a theoretical discussion of optimization concepts and procedures, and implicitly assumes that optimization enhances a system's *future* performance. As discussed in the next section, however, the viability of optimization is open to serious question.

The Optimization Myth

It is ironic that optimization receives so much attention while its underlying premise is rarely considered. In other words, do the better performing parameter sets of the *past* continue to exhibit above-average performance in the *future?*

As an empirical test of the validity of optimization we examine the historical rankings of a range of parameter set values for a breakout system: reverse from short to long if today's close is higher than the highest close during the past N days; reverse from long to short if today's close is lower than the lowest close during the past N days. Nine values of N for this system were tested: 20, 30, 40, 50, 60, 70, 80, 90, and 100.

Tables 19.1 to 19.10 compare the profit/loss rankings of these parameter sets in 10 markets for three 2-year test periods (2009–2010, 2011–2012, and 2013–2014), with parameter sets listed in the order of their performance during the respective *prior* eight-year periods. (All markets were traded with one contract per signal.) In other words, the top-performing parameter set of the prior eight-year period (2001–2008, 2003–2010, or 2005–2012) is listed first, the second-best parameter set of the prior period is listed second, and so on. For example, if the top number in a column is 6, it means that the best-performing parameter set for that market in the prior eight-year period was the sixth-ranked parameter set (out of nine) during the given test period.

TABLE 19.1 Breakout System (10-Year T-Notes): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	9	9	7
2	8	6	5
3	7	7	3
4	2	8	1
5	5	4	4
6	6	5	6
7	1	3	2
8	3	1	9
9	4	2	8

TABLE 19.2 Breakout System (Euro): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	4	2	1
2	9	1	7
3	5	4	2
4	6	5	5
5	7	6	8
6	3	3	3
7	8	7	9
8	2	8	6
9	1	9	4

TABLE 19.3 Breakout System (Japanese Yen): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	9	5	4
2	2	3	1
3	8	7	6
4	1	6	2
5	3	1	7
6	4	4	8
7	7	9	9
8	6	2	5
9	5	8	3

TABLE 19.4 Breakout System (Gold): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	7	2	2
2	3	4	3
3	4	5	4
4	9	1	9
5	6	6	1
6	8	9	7
7	1	8	5
8	2	7	6
9	5	3	8

TABLE 19.5 Breakout System (Natural Gas): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	8	3	1
2	4	5	4
3	5	1	2
4	1	6	3
5	6	8	9
6	2	9	5
7	9	4	8
8	7	7	6
9	3	2	7

TABLE 19.6 Breakout System (WTI Crude Oil): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	3	6	1
2	2	7	6
3	7	9	8
4	4	1	2
5	5	3	5
6	1	5	4
7	9	8	9
8	6	2	3
9	8	4	7

TABLE 19.7 Breakout System (Corn): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	3	7	3
2	4	1	7
3	2	3	5
4	1	8	8
5	9	4	1
6	5	9	6
7	6	2	2
8	8	5	4
9	7	6	9

TABLE 19.8 Breakout System (Soybeans): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	6	4	5
2	3	5	3
3	4	7	1
4	1	2	4
5	2	3	2
6	8	1	7
7	7	6	6
8	9	8	8
9	5	9	9

TABLE 19.9 Breakout System (Coffee): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	3	1	9
2	8	2	1
3	1	6	6
4	7	8	3
5	9	9	2
6	2	5	4
7	6	7	8
8	5	4	7
9	4	3	5

TABLE 19.10 Breakout System (E-Mini Nasdaq 100): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 2009–2010	Rank of Same Parameter Set in 2011–2012	Rank of Same Parameter Set in 2013–2014
1	5	3	9
2	7	1	7
3	4	2	8
4	2	8	4
5	6	6	6
6	9	5	1
7	3	9	2
8	8	4	5
9	1	7	3

As a visual aid to help see if there is any consistency between past and future performance, the two top-performing parameter sets in each test period are denoted by circles and the two bottom parameter sets by squares. If the basic premise of optimization were valid—that is, that the best-performing parameter sets of the *past were* likely to be the best-performing parameter sets in the *future*—then Tables 19.1 through 19.10 should reflect a pattern of circles consistently near column tops and squares consistently near column bottoms. However, this is not the case. Both circles and squares are sometimes near column tops, sometimes near column bottoms, and sometimes near column midpoints. The apparent randomness in the vertical placement of the circles and squares in Tables 19.1 through 19.10 implies the correlation between past and future performance is highly tenuous.

Table 19.11 further highlights the weakness of the relationship between past and future performance. In addition to showing the average rank of the best-performing parameter sets from the eight-year sample periods in the subsequent two-year test periods (second column), Table 19.11 also shows how often the best- and worst-performing sets in a prior eight-year period repeated their positions in the subsequent two-year period versus completely reversing their rank order. Note the initially best- and worst-performing parameter sets repeated in subsequent two-year periods a total of eight times, which is only one time more than the number of times the best set became the worst set or the worst set became the best set. Also notice that the best-performing parameter set became the worst-performing set one more time (5) than the best-performing set repeated as the top set.

This instability in the values of the best-performing parameter sets from period to period means gauging a system's performance by the best *past* parameter sets will grossly overstate the system's performance potential. To illustrate this point, Tables 19.12 through 19.15 compare the performance of the best parameter set in each test period versus the average of all parameter sets and the performance of the parameter sets that had the best and worst results in the

TABLE 19.11 Stability of Best- and Worst-Performing Parameter Sets

Market	Avg. Rank of Best Parameter Set	Best Parameter Set Repeated	Worst Parameter Set Repeated	Best Set Becomes Worst Set	Worst Set Becomes Best Set
10-yr. T-note	4.70	0	0	2	0
Euro	4.30	1	1	0	1
Japanese yen	4.77	0	0	1	0
Gold	4.27	0	0	0	0
Natural gas	5.10	1	0	0	0
WTI crude oil	5.13	1	0	0	0
Corn	6.00	0	1	0	0
Soybeans	5.43	0	2	0	0
Coffee	5.30	1	0	1	0
E-mini Nasdaq 100	4.70	0	0	1	1
Total		**4**	**4**	**5**	**2**

prior period. In this example, based on the all-market totals, selecting the worst parameter set in the prior period would have outperformed a strategy of picking the best past parameter set in one of the three test periods (see Table 19.12), as well as the three-period total (see Table 19.15). The penultimate column of these tables marks the instances the worst-performing parameter set in a prior eight-year period outperformed the prior best-performing set in the subsequent two-year period. The final column shows how often the average parameter set performance in the subsequent two-year period outperformed the best-performing set of the prior eight-year period.

TABLE 19.12 Profit/Loss ($) Comparisons for 2009–2010 Test Period: Actual Best Parameter Set vs. Period Average and Best and Worst Parameter Sets in Prior Period

Market	Best Parameter Set in Period	Best Parameter Set in Prior Period	Worst Parameter Set in Prior Period	Avg. of all Parameter Sets	Worst Prior > Best Prior	Avg. > Best Prior
10-yr. T-note	$7,453	–$7,188	$2,391	$253	X	X
Euro	$47,575	$18,963	$47,575	$22,511	X	X
Japanese yen	$5,438	–$23,825	–$9,638	–$8,967	X	X
Gold	$50,740	$7,420	$19,020	$25,084	X	X
Natural gas	$46,960	–$7,360	$34,120	$16,522	X	X
WTI crude oil	–$11,670	–$26,030	–$45,150	–$33,041		
Corn	$8,875	$6,913	–$338	$3,188		
Soybeans	$34,188	$11,875	$22,350	$16,944	X	X
Coffee	$25,650	$12,075	$11,963	$6,713		
E-Mini Nasdaq 100	$12,330	$4,820	$12,330	$5,417	X	X
Total	**$227,538**	**–$2,338**	**$94,623**	**$54,625**	**7**	**7**

TABLE 19.13 Profit/Loss ($) Comparisons for 2011–2012 Test Period: Actual Best Parameter Set vs. Period Average and Best and Worst Parameter Sets in Prior Period

Market	Best Parameter Set in Period	Best Parameter Set in Prior Period	Worst Parameter Set in Prior Period	Avg. of All Parameter Sets	Worst Prior > Best Prior	Avg. > Best Prior
10-yr. T-note	$13,172	–$3,750	$9,234	$3,516	X	X
Euro	$10,900	$10,900	–$11,550	$1,938		
Japanese yen	–$1,538	–$7,963	–$12,913	–$8,157		
Gold	$16,310	$7,300	$3,170	–$5,672		
Natural gas	$16,050	$2,590	$10,930	–$712	X	
WTI crude oil	$12,330	–$30,950	–$11,920	–$19,537	X	X
Corn	–$963	–$8,563	–$8,538	–$9,138	X	
Soybeans	$24,013	–$3,113	–$16,413	–$2,590		X
Coffee	$48,563	$48,563	$20,963	$8,308		
E-Mini Nasdaq 100	$1,540	–$7,630	–$20,870	–$13,506		
Total	**$140,377**	**$7,385**	**–$37,906**	**–$45,550**	**4**	**3**

TABLE 19.14 **Profit/Loss ($) Comparisons for 2013–2014 Test Period: Actual Best Parameter Set vs. Period Average and Best and Worst Parameter Sets in Prior Period**

Market	Best Parameter Set in Period	Best Parameter Set in Prior Period	Worst Parameter Set in Prior Period	Avg. of All Parameter Sets	Worst Prior > Best Prior	Avg. > Best Prior
10-yr. T-note	$2,922	–$2,328	–$3,359	–$1,557		X
Euro	$19,963	$19,963	$5,013	$2,568		
Japanese yen	$39,713	$38,138	$39,713	$27,339	X	
Gold	$25,840	$21,160	–$4,340	$11,042		
Natural gas	$6,250	$6,250	–$1,590	–$2,077		
WTI crude oil	$39,060	$39,060	$18,070	$23,379		
Corn	$9,750	$3,675	–$1,150	$2,661		
Soybeans	$8,663	$488	–$12,863	$1,211		X
Coffee	$28,313	–$9,113	$2,963	$7,677	X	X
E-Mini Nasdaq 100	$29,640	–$8,780	$16,505	$10,635	X	X
Total	**$210,112**	**$108,512**	**$58,961**	**$82,878**	**3**	**4**

TABLE 19.15 **Profit/Loss ($) Comparisons for Three Test Periods Combined: Actual Best Parameter Sets vs. Period Averages and Best and Worst Parameter Sets in Prior Periods**

Market	Best Parameter Set in Period Total	Best Parameter Set in Prior Period Total	Worst Parameter Set in Prior Period Total	Avg. of All Parameter Sets Total	Worst Prior > Best Prior	Avg. > Best Prior
10-yr. T-note	$23,547	–$13,266	$8,266	$2,212	X	X
Euro	$78,438	$49,825	$41,038	$27,017		
Japanese yen	$43,613	$6,350	$17,163	$10,215	X	X
Gold	$92,890	$35,880	$17,850	$30,454		
Natural gas	$69,260	$1,480	$43,460	$13,733	X	X
WTI crude oil	$39,720	–$17,920	–$39,000	–$29,199		
Corn	$17,663	$2,025	–$10,025	–$3,289		
Soybeans	$66,863	$9,250	–$6,925	$15,565		X
Coffee	$102,525	$51,525	$35,888	$22,698		
E-Mini Nasdaq 100	$43,510	–$11,590	$7,965	$2,546	X	X
Total	**$578,027**	**$113,559**	**$115,678**	**$91,953**	**4**	**5**

Our example used a very small list of only nine parameter sets. Many system developers run optimizations across hundreds or even thousands of parameter sets. Imagine the degree of performance overstatement that would occur by representing a system's performance by the best parameter sets in these cases!

For comparison, Tables 19.16 through 19.19 show the same information as Tables 19.12 through 19.15 except they reflect tests of the same system conducted 20 years earlier on a slightly different portfolio (30-year U.S. T-bonds, Deutsche marks, Japanese yen, gold, silver, heating oil, corn, soybeans, live cattle, and sugar). In this case, the three 8-year sample periods were 1981–1988, 1983–1990, and 1985–1992 and the three 2-year test periods were 1989–1990, 1991–1992, and 1993–1994.

TABLE 19.16 **Profit/Loss ($) Comparisons for 1989–1990 Test Period: Actual Best Parameter Set vs. Period Average and Best and Worst Parameter Sets in Prior Period**

Market	Best Parameter Set in Period	Best Parameter Set in Prior Period	Worst Parameter Set in Prior Period	Avg. of All Parameter Sets	Worst Prior > Best Prior	Avg. > Best Prior
T-bond	6,670	−9,090	1,420	−2,180	X	X
Deutsche mark	7,780	3,020	6,340	5,390	X	X
Japanese yen	11,840	9,240	8,420	8,130		
Gold	3,390	1,700	−320	1,080		
Silver	5,850	5,330	1,630	3,050		
Heating oil	7,650	1,760	6,430	3,380	X	X
Corn	1,640	−2,190	−2,730	−590		X
Soybeans	4,970	−7,160	4,740	−740	X	X
Cattle	2,090	850	−3,290	−20		
Sugar	4,240	4,170	−5,560	−840		
Total	**56,120**	**7,630**	**17,080**	**16,030**	**4**	**5**

TABLE 19.17 **Profit/Loss ($) Comparisons for 1991–1992 Test Period: Actual Best Parameter Set vs. Period Average and Best and Worst Parameter Sets in Prior Period**

Market	Best Parameter Set in Period	Best Parameter Set in Prior Period	Worst Parameter Set in Prior Period	Avg. of All Parameter Sets	Worst Prior > Best Prior	Avg. > Best Prior
T-bond	3,710	−1,820	−2,920	−420		X
Deutsche mark	9,180	1,680	9,180	4,770	X	X
Japanese yen	3,340	−240	−3,620	−1,670		
Gold	1,370	90	1,370	−1,050	X	
Silver	−720	−1,890	−1,780	−1,640	X	X
Heating oil	5,510	−980	4,290	1,540	X	X
Corn	560	−480	340	−440	X	X
Soybeans	−2,420	−6,090	−3,190	−4,650	X	X
Cattle	1,380	−160	1,380	−340	X	
Sugar	810	−1,690	−1,850	−1,410		X
Total	**22,700**	**−11,570**	**3,200**	**−5,010**	**7**	**7**

TABLE 19.18 **Profit/Loss ($) Comparisons for 1993–1994 Test Period: Actual Best Parameter Set vs. Period Average and Best and Worst Parameter Sets in Prior Period**

Market	Best Parameter Set in Period	Best Parameter Set in Prior Period	Worst Parameter Set in Prior Period	Avg. of All Parameter Sets	Worst Prior > Best Prior	Avg. > Best Prior
T-bond	11,600	3,500	7,910	7,180	X	X
Deutsche mark	6,210	−3,660	−1,410	−3,300	X	X
Japanese yen	3,620	2,460	−3,060	260		
Gold	490	−1,900	−930	−1,460	X	X
Silver	1,600	−3,650	−790	−2,690	X	X
Heating oil	2,200	2,200	−890	−1,700		
Corn	1,910	1,910	−1,030	640		
Soybeans	2,120	1,570	−2,060	−240		
Cattle	1,600	950	1,600	500	X	
Sugar	880	570	−240	−550		
Total	**32,230**	**3,950**	**−900**	**−1,360**	**5**	4

TABLE 19.19 **Profit/Loss ($) Comparisons for Three Test Periods Combined: Actual Best Parameter Sets vs. Period Averages and Best and Worst Parameter Sets in Prior Periods**

Market	Best Parameter Sets in Test Periods Total	Best Parameter Sets in Prior Periods Total	Worst Parameter Sets in *Prior* Periods Total	Period Parameter Set Averages Total	Worst Prior > Best Prior	Avg. > Best Prior
T-bond	21,980	−7,410	6,410	3,950	X	X
Deutsche mark	23,170	1,040	14,110	6,860	X	X
Japanese yen	18,800	11,460	1,740	6,720		
Gold	5,250	−110	120	−1,430	X	
Silver	6,730	−210	−940	−1,280		
Heating oil	15,360	2,980	9,830	3,220	X	X
Corn	4,110	−760	−3,420	−390		X
Soybeans	4,670	−11,680	−510	−5,330	X	X
Cattle	5,070	1,640	−310	140		
Sugar	5,930	3,060	−7,650	−2,800		
Total	**111,070**	**10**	**19,380**	**9,660**	**5**	**5**

Based on the combined three-period, all-market totals from this second set of tests, selecting the worst parameter set in the prior period actually would have outperformed a strategy of picking the best past parameter set in two of the three test periods, as well as the three-period total!

This observation is not intended to imply that the prior-period worst-performing parameter set is likely to outperform the prior-period best-performing set. If similar empirical tests were conducted for other systems, the prior-period best-performing parameter set would probably outperform the prior-period worst-performing set more often than the other way around (although the types of results witnessed in our example are far from uncommon). The key point, however, is that invariably, as was the case in Tables 19.12 through 19.15 and 19.16 through 19.19, the prior-period best-performing parameter sets would fall far short of the actual best-performing parameter sets for the given periods and would often fail to provide any statistically significant improvement over the average of all parameter sets.

Although optimization seemed to have little, if any, value when applied market by market, optimization does appear to be a bit more useful if applied to a portfolio. In other words, instead of picking the best past parameter set for each market, the best past single parameter set applied across all markets is selected. Table 19.20 shows the two-year test period parameter set rankings for a portfolio consisting of the 10 markets that provided the results for Tables 19.16 through 19.19.[6] The one striking correlation between past and future performance is that the worst parameter set in the prior eight-year period is also the worst parameter set in the subsequent two-year period in all three test intervals!

Although the worst past parameter set also seems likely to be the worst future parameter set, other past ranking placements seem to imply little predictive value. The average ranking for all three test periods of the remaining eight prior-period ranking placements (i.e., all rankings excluding the worst one) is 4.5. While the average test period ranking of the best parameter set in the prior eight-year period (3.3) is somewhat better than this average, the fourth-ranked parameter set in the prior period has by far the best average ranking in the future test periods (2.3). Also note that the second-best prior-period parameter set has an average test period rank almost identical to the corresponding average for the second worst prior-period parameter set (4.7 vs. 5.0).

To gain some insight as to why the worst prior-period ranking seems to be such an excellent predictor of future performance (namely, continued poor performance for that parameter set), while other ranking placements seem to have little predictive value, we examine performance rankings based on parameter set value. Table 19.21 indicates parameter set rankings in each of the three tests periods based on parameter set values (as opposed to prior-period rankings as was the case in Table 19.20). The parameter set values are listed in ascending order.

TABLE 19.20 Breakout System (Portfolio): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods

Parameter Set Rank Prior Eight-Year Period	Rank of Same Parameter Set in 1989–1990	Rank of Same Parameter Set in 1991–1992	Rank of Same Parameter Set in 1993–1994	Avg. Rank
1	1	7	2	3.3
2	5	1	8	4.7
3	3	6	4	4.3
4	2	4	1	2.3
5	4	8	6	6.0
6	6	3	7	5.3
7	7	5	3	5.0
8	8	2	5	5.0
9	9	9	9	9.0

[6] In this case the portfolio consisted of one contract in each market, with the exception of corn, which was traded with two contracts because of its low volatility.

TABLE 19.21 Breakout System (Portfolio): Comparison of Parameter Set Rankings in Two-Year Test Periods Based on *N*-Values

Parameter Set *N*-Value	Rank of Parameter Set in 1989–1990	Rank of Parameter Set in 1991–1992	Rank of Parameter Set in 1993–1994	Avg. Rank
20	9	9	9	9.0
30	8	2	5	5.0
40	7	5	3	5.0
50	6	3	1	3.3
60	4	6	6	5.3
70	5	7	8	6.7
80	1	1	2	1.3
90	2	4	4	3.3
100	3	8	7	6.0

Table 19.21 reveals that the worst-performing parameter set in each of the test periods was actually the same parameter set! (Since Table 19.20 indicated the test period worst parameter set was the same as the prior-period worst parameter set in all three cases, the implication is that this same parameter set was also the worst-performing parameter set in all three prior eight-year periods.) This consistently worst-performing parameter set is at one extreme end of the parameter set range tested: $N = 20$.

Although $N = 20$—the most sensitive parameter set value tested—is consistently the worst performer (when applied across a portfolio), the other values tested ($N = 30$ to $N = 100$) show no consistent pattern. It is true that the parameter set $N = 80$ was by far the best-performing set with an incredible average rank of 1.3. However, the average rankings of the two surrounding N-values (6.7 and 3.3) suggest that the stellar performance of $N = 80$ was probably a statistical fluke. As was explained earlier in this chapter, a lack of *parameter stability* suggests that the past superior performance of a parameter set probably reflects a peculiarity in the historical data tested rather than a pattern that is likely to be repeated in the future.

Tables 19.22 and 19.23 show analogous portfolio optimization statistics for the portfolios in the more recent test periods that were reviewed in Tables 19.12 through 19.15. Note in Table 19.23 that the same $N = 20$ parameter set once again exhibited inferior performance, registering as the worst-performing set in two of the three periods.

It is instructive to review the observations revealed by the foregoing optimization experiments:

- Optimization appeared to have no value whatsoever when applied on a market-by-market basis.
- When applied to a portfolio, however, optimization in the earlier (1981–1994) example appeared useful in predicting the parameter set most likely to witness inferior future performance, although it still showed no reliable pattern in predicting the parameter set most likely to witness superior future performance.
- Upon closer examination it appeared this pattern of consistent inferior performance was not so much a consequence of the prior-period ranking as the parameter value. In other words, the

TABLE 19.22 Breakout System (Portfolio): Comparison of Parameter Set Rankings in Two-Year Test Periods vs. Rankings in Prior Eight-Year Periods (2000–2014)

Parameter Set Rank Prior Eight-Year Period	Parameter Set Rank in 2009–2010	Parameter Set Rank in 2011–2012	Parameter Set Rank in 2013–2014	Average Rank
1	9	1	1	3.7
2	2	7	5	4.7
3	3	2	3	2.7
4	7	3	9	6.3
5	6	4	4	4.7
6	8	5	8	7.0
7	4	8	2	4.7
8	5	6	7	6.0
9	1	9	6	5.3

TABLE 19.23 Breakout System (Portfolio): Comparison of Parameter Set Rankings in Two-Year Test Periods Based on *N*-Values (2000–2014)

Parameter Set *N*-Value	Rank of Parameter Set in 2009–2010	Rank of Parameter Set in 2011–2012	Rank of Parameter Set in 2013–2014	Average Rank
20	9	3	9	7.0
30	7	5	8	6.7
40	6	8	7	7.0
50	8	9	6	7.7
60	5	6	2	4.3
70	3	7	3	4.3
80	4	4	4	4.0
90	1	2	5	2.7
100	2	1	1	1.3

parameter set range tested began at a value that was clearly suboptimal for the given system: $N = 20$. This same parameter value remained suboptimal, on average, in the more recent test period as well. Although not indicated in the parameter set ranking tables, lower values of N would have shown even worse performance—in fact, strikingly worse—as the value of N was decreased.

These observations, which are consistent with the results of other similar empirical tests I have conducted in the past, suggest the following five key conclusions regarding optimization:[7]

[7] Although a single empirical experiment cannot be used to draw broad generalizations, I am willing to do so here because the results of the optimization test just described are fairly typical of many similar tests I have conducted in the past. In this sense, the optimization tests detailed in the text are not intended as a *proof* of the severe limitations of optimization, but rather as *illustrations* of this point.

1. Any system—repeat, any system—can be made to be very profitable through optimization (i.e., over its past performance). If you ever find a system that can't be optimized to show good profits in the past, congratulations, you have just discovered a money machine (by doing the opposite, unless transaction costs are exorbitant). Therefore, a wonderful past performance for a system that has been optimized may be nice to look at, but it doesn't mean very much.
2. Optimization will always, repeat always, overstate the potential future performance of a system—usually by a wide margin (say, three trailer trucks' worth). Therefore, optimized results should never, repeat never, be used to evaluate a system's merit.
3. For many if not most systems, optimization will improve *future* performance only marginally, if at all.
4. If optimization has any value, it is usually in defining the broad boundaries for the ranges from which parameter set values in the system should be chosen. Fine-tuning of optimization is at best a waste of time and at worst self-delusion.
5. In view of the preceding items, sophisticated and complex optimization procedures are a waste of time. The simplest optimization procedure will provide as much meaningful information (assuming that there is any meaningful information to be derived).

In summary, contrary to widespread belief, there is some reasonable question as to whether optimization will yield meaningfully better results over the long run than randomly picking the parameter sets to be traded. Lest there be any confusion, let me explicitly state that this statement is not intended to imply that optimization is never of any value. First, as indicated previously, optimization can be useful in defining the suboptimal extreme ranges that should be excluded from the selection of parameter set values (e.g., $N \leq 20$ in our breakout system example). Also, it is possible that, for some systems, optimization may provide some edge in parameter set selection, even after suboptimal extreme ranges are excluded. However, I do mean to imply that the degree of improvement provided by optimization is far less than generally perceived and that traders would probably save a lot of money by first proving any assumptions they are making about optimization rather than taking such assumptions on blind faith.

Testing versus Fitting

Perhaps the most critical error made by users of futures trading systems is the assumption the performance of the optimized parameter sets during the test period provides an approximation of the potential performance of those sets in the future. As was demonstrated in the previous section, such assumptions will lead to grossly overstated evaluations of a system's true potential. It must be understood that futures market price fluctuations are subject to a great deal of randomness. Thus, the "ugly truth" is that the question of which parameter sets will perform best during any given period is largely a matter of chance. The laws of probability indicate that if enough parameter sets are tested, even a meaningless trading system will yield some sets with favorable past performance. Evaluating a system based on the optimized parameter sets (i.e., the best-performing sets during the survey period) would be best described as fitting the system to past results rather than testing the system. If optimization can't be used to gauge performance, how then do you evaluate a system? The following sections describe two meaningful approaches.

Blind Simulation

In the blind simulation approach the system is optimized using data for a time period that deliberately excludes the most recent years. The performance of the system is then tested using the selected parameter sets for subsequent years. Ideally, this process should be repeated several times.

Note that the error of fitting results is avoided because the parameter sets used to measure performance in any given period are selected entirely on the basis of prior rather than concurrent data. In a sense, this testing approach mimics real life (i.e., one must decide which parameter sets to trade on the basis of past data).

The optimization tests of the previous section used this type of procedure, stepping through time in two-year intervals. Specifically, system results for the 2001–2008 period were used to select the best-performing parameter sets, which were then tested for the 2009–2010 period. Next, the system results for the 2003–2010 period were used to select the best-performing parameter sets, which were then tested for the 2011–2012 period. Finally, the system results for the 2005–2012 period were used to select the best-performing parameter sets, which were then tested for the 2013–2014 period.

The essential point is that simulation and optimization periods should not be allowed to overlap. Simulations that are run over the same period as the optimization are worthless.

Average Parameter Set Performance

Finding the average parameter set performance requires defining a complete list of all parameter sets you wish to test *before* running any simulations. Simulations are then run for all the parameter sets selected, and the average of all sets tested is used as an indication of the system's potential performance. This approach is valid because you could always throw a dart to pick a parameter from a broad range of parameter set values. If you throw enough darts, the net result will be the average. The important point is that this average should be calculated across all parameter sets, not just those sets that prove profitable. Note that the trader might still choose to trade the optimized parameter sets for the future (instead of randomly selected ones), but the evaluation of the system's performance should be based on the average of all sets tested (which is equivalent to a random selection process).

The blind simulation approach probably comes closest to duplicating real-life trading circumstances. However, the average parameter set performance is probably as conservative and has the advantage of requiring far less calculation. Both approaches represent valid procedures for testing a system.

One important caveat: In the advertised claims for given systems, the term *simulated results* is often used loosely as a euphemism for optimized results (instead of implying the results are based on a blind simulation process). If this is the case, the weight attached to the results should equal the amount of money invested in the system: zero. The commonplace misuse and distortion of simulated results is examined in detail in the next section.

The Truth about Simulated Results

Although the value of optimization in improving a system's future performance is open to debate, there is absolutely no question the use of optimized results will greatly distort the implied future performance of a system. As was demonstrated earlier in this chapter, there is very little, if any, correlation between the best-performing parameters in a system for one period and the best-performing parameters in a subsequent period. Hence, assuming that the performance implied by the best-performing parameters could have been achieved in the past is totally unrealistic.

After years of experience, my attitude toward simulated results is summarized by what I call Schwager's simulations corollary to Gresham's law of money. As readers may recall from Economics 101, Gresham's proposition was that "bad money drives out good." Gresham's contention was that if two types of money were in circulation (e.g., gold and silver) at some arbitrarily defined ratio (e.g., 16:1), the bad money (i.e., the money overvalued at the fixed rate of exchange) would drive out the good. Thus, if gold were worth more than 16 ounces of silver, a 16:1 ratio would result in silver driving gold out of circulation (as people would tend to hoard it).

My corollary is "bad simulations drive out good." The term *bad* means simulations derived based on highly tenuous assumptions, not bad in terms of indicated performance. On the contrary, truly "bad" simulations will show eye-popping results.

I frequently see ads hawking systems that supposedly make 200 percent, 400 percent, or even 600 percent a year. Let's be conservative—and I use the term loosely—and assume a return of *only* 100 percent per year. At this level of return, $100,000 would grow to over $1 *billion* in just over 13 years! How can such claims possibly be true, then? The answer is they can't. The point is that, given enough hindsight, it is possible to construct virtually any type of past-performance results. If anyone tried to sell a system or a trading program based on truly realistic simulations, the results would appear laughably puny relative to the normal promotional fare. It is in this sense that I believe that bad (unrealistic) simulations drive out good (realistic) simulations.

How are simulated results distorted? Let us count the ways:

1. **The well-chosen example (revisited).** In constructing a well-chosen example, the system promoter selects the best market, in the best time period, using the best parameter set. Assuming a system is tested on 25 markets for 15 years and uses 100 parameter set variations, there would be a total of 37,500 (25 × 15 × 100) one-year results. It would be difficult to construct a system in which not one of these 37,500 possible outcomes showed superlative results. For example, if you tossed a group of 10 coins 37,500 times, don't you think you would get 10 out of 10 heads sometimes? Absolutely. In fact, you would get 10 out of 10 heads on the average of one out of 1,024 times.
2. **Kitchen sink approach.** By using hindsight to add parameters and create additional system rules that conveniently take care of past losing periods, it is possible to generate virtually any level of past performance.
3. **Ignoring risk.** Advertised system results frequently calculate return as a percent of margin or as a percent of an unrealistically low multiple of margin. This return measurement approach

alone can multiply the implied returns severalfold. Of course, the risk would increase commensurately, but the ads don't provide those details.

4. **Overlooking losing trades.** It is hardly uncommon for charts in system websites or advertisements to indicate buy and sell signals at the points at which some specified rules were met, but fail to indicate other points on the same chart where the same conditions were met and the resulting trades were losers.
5. **Optimize, optimize, optimize.** Optimization (i.e., selecting the best-performing parameter sets for the *past*) can tremendously magnify the past performance of a system. Virtually any system ever conceived by man would look great if the results were based on the best parameter set (i.e., the parameter set that had the best past performance) for each market. The more parameter sets tested, the wider the selection of past results, and the greater the potential simulated return.
6. **Unrealistic transaction costs.** Frequently, simulated results only include commissions but not slippage (the difference between the assumed entry level and the actual fill that would be realized by using a market or stop order). For short-term systems (e.g., those using intraday data), ignoring slippage can make a system that would wipe out an account in real life look like a money machine.
7. **Fabrication.** Even though it is remarkably easy to construct system rules with great performance for the past, some promoters don't even bother doing this much. For example, one infamous individual for years repeatedly promoted $299 systems that were outright frauds.

The preceding is not intended to indict all system promoters or those using simulated results. Certainly, there are many individuals who construct simulated results in appropriately rigorous fashion. However, the sad truth is that the extraordinary misuse of simulations over many years has made simulated results virtually worthless. Advertised simulated results are very much like restaurant reviews written by the proprietors—you would hardly expect to ever see a bad review. I can assure you that you will never see any simulated results for a system that show the system long the S&P as of the close of October 16, 1987, September 10, 2001, or March 5, 2010. Can simulated results ever be used? Yes, if you are the system developer *and* you know what you're doing (e.g., use the simulation methods detailed in the previous section), or, equivalently, if you have absolute faith in the integrity and competence of the system developer.

Multimarket System Testing

Although it is probably unrealistic to expect any single system to work in all markets, generally speaking, a good system should demonstrate profitability in a large majority of actively traded markets (e.g., 85 percent or more). There are, of course, some important exceptions. A system employing fundamental input would, by definition, be applicable to only a single market. In addition, the behavior of some markets is so atypical (e.g., stock indexes) that systems designed for trading such markets might well perform poorly over the broad range of markets.

In testing a system for a multimarket portfolio, it is necessary to predetermine the relative number of contracts to be traded in each market. This problem is frequently handled by simply assuming the system will trade one contract in each market. However, this is a rather naive approach, for two reasons. First, some markets are far more volatile than other markets. For example, a portfolio that included one contract of coffee and one contract of corn would be far more dependent on the trading results in coffee. Second, it may be desirable to downgrade the relative weightings of some markets because they are highly correlated with other markets (e.g., 10-year T-notes and 30-year T-bonds).[8]

In any case, the percentage allocation of available funds to each market should be determined prior to testing a system. These relative weightings can then be used to establish the number of contracts to be traded in each market.

Negative Results

One should not overlook the potential value of negative results. Analyzing the conditions under which a system performs poorly can sometimes reveal important weaknesses in the system that have been overlooked and thus provide clues as to how the system can be improved. Of course, the fact that the implied rule changes improve results in the poorly performing case does not prove anything. However, the validity of any suggested rule changes would be confirmed if such revisions generally tended to improve the results for other parameter sets and markets as well. The potential value of negative results as a source of ideas for how a system can be improved cannot be overstated. The concept that disorder is a catalyst for thought is a general truth that was perfectly expressed by the late novelist John Gardner: "In a perfect world, there would be no need for thought. We think because something goes wrong."

The idea of learning from poor results is basically applicable to a system that works in most markets and for most parameter sets but performs badly in isolated cases. However, systems that exhibit disappointing results over a broad range of markets and parameter sets are likely to be lost causes, unless the results are spectacularly poor. In the latter case, a system that exactly reverses the trade signals of the original system might be attractive. For example, if tests of a new trend-following system reveal that the system consistently loses money in most markets, the implication is that one might have accidently stumbled upon an effective countertrend system. Such discoveries may be difficult on the ego, but they should not be ignored.

Of course, the fact that a system exhibits stable poor performance does not imply that the reverse system would perform favorably, since transaction costs may account for a significant portion of losses. Thus, the reverse system might also perform badly once these costs are taken into account, as was the case for the aforementioned well-chosen example described at the start of

[8] For purposes of future trading (as opposed to historical testing), historical performance might be a third relevant factor in determining contract weightings. However, this factor cannot be included as an input in the testing procedure because it would bias the results.

this chapter. As another example, at surface glance, reversing the signals generated by a system that loses an average of $3,000 per year may appear to be an attractive strategy. If, however, two-thirds of the loss can be attributed to transaction costs, fading the signals of this system will result in a loss of $1,000 per year, assuming a continuation of the same performance. (The preceding assumptions imply that transaction costs equal $2,000 per year and that the trades lose $1,000 per year net of these costs. Thus, reversing the signals would imply a $1,000-per-year gain on the trades, but the $2,000-per year transaction costs would imply a net loss of $1,000 per year.) Moral: If you are going to design a bad system, it should be truly terrible if it is to be of value.

Ten Steps in Constructing and Testing a Trading System

1. Obtain all data needed for testing. Again, with the exception of short-term trading systems, which may be able to use actual contract data, the use of continuous futures (not to be confused with nearest futures or perpetual prices) is highly recommended.
2. Define the system concept.
3. Program rules to generate trades in accordance with this concept.
4. Select a small subset of markets and a subset of years for these markets.
5. Generate system trading signals for this subset of markets and time for a given parameter set.
6. Check to see that the system is doing what was intended. Almost invariably, a careful check will reveal some inconsistencies due to either or both of the following reasons:
   a. There are errors in the program.
   b. Rules in program do not anticipate some circumstances, or they create unforeseen repercussions.

   Some examples of the latter might include the system failing to generate a signal, given an event at which a signal is intended; system generating a signal when no signal is intended; system rules inadvertently creating a situation in which no new signals can be generated or in which a position is held indefinitely. In essence, these types of situations arise because there will often be some missed nuances.

   The system rules need to be modified to correct both programming errors as well as unforeseen inconsistencies. It should be emphasized that corrections of the latter type are only concerned with making the system operate consistently with the intended concept and should be made *without any regard as to whether the changes help or hurt performance in the sample cases used in the developmental process.*
7. After making necessary corrections, repeat step 6. Pay particular attention to changes in the indicated signals versus those from previous runs for two reasons:
   a. To check whether the program changes achieved the desired fix.
   b. To make sure the changes did not have unintended effects.
8. Once the system is working as intended, and all rules and contingencies have been fully defined, *and only after such a point,* test the system on the entire defined parameter set list across the full database. Be sure the intended trading portfolio has been defined before this test is run.

9. As detailed earlier in this chapter, evaluate performance based on the average of all parameter sets tested or a blind simulation process. (The former involves far less work.)
10. Compare these results with the results of a generic system (e.g., breakout, crossover moving average) for the corresponding portfolio and test period. The return/risk of the system should be *measurably* better than that of the generic system if it is to be deemed to have any real value.

The preceding steps represent a rigorous procedure that is designed to avoid generating results that are upwardly biased by hindsight. As such, expect most system ideas to fail the test of merit in step 10. Designing a system with a truly superior performance is more difficult than most people think.

Observations about Trading Systems

1. In trend-following systems, the basic method used to identify trends (e.g., breakout, crossover moving average) may well be the least important component of the system. In a sense, this contention is merely a restatement of Jim Orcutt's observation that "There are only two types of trend-following systems: fast and slow." Thus, in designing trend-following systems, it may make more sense to concentrate on modifications (e.g., filters and confirmation rules to reduce bad trades, market characteristic adjustments, pyramiding rules, stop rules) than on trying to discover a better method for defining trends.
2. Complexity for its own sake is no virtue. Use the simplest form of a system that does not imply a meaningful sacrifice in performance relative to more complex versions.
3. The well-publicized and very valid reason for trading a broad range of markets is risk control through diversification. However, there is a very important additional reason for trading as many markets as possible: insurance against missing any of the sporadic giant price moves in the futures markets. The importance of catching all such major trends cannot be overstressed—it can make the difference between mediocre performance and great performance. The 2008–2011 gold market and the 2007–2009 and 2014–2016 crude oil markets are three spectacular examples of markets that were critical to portfolio performance.
4. If trading funds are sufficient, diversification should be extended to systems as well as markets. Trading several systems rather than a single system could help smooth overall performance. Ideally, the greatest degree of diversification would be achieved if the mix of systems included countertrend and pattern-recognition systems as well as trend-following systems. (However, this goal may be difficult to achieve because countertrend and pattern-recognition systems are generally significantly harder to design than trend-following systems.)
5. If sufficient funds are available, it is better to trade a number of diversified parameter sets than to trade a single optimized set.
6. Generally speaking, the value of parameter optimization is far overstated.
7. The previous observation strongly suggests that optimized results should never be used for evaluating the relative performance of a system. Two meaningful methods for testing systems were discussed in the text.

8. So-called *simulated* results are frequently *optimized* results (i.e., derived with the benefit of hindsight) and, as such, virtually meaningless. This caveat is particularly pertinent in regard to promotions for trading systems, which invariably use very well-chosen examples.
9. An analysis of the results of successful systems will almost invariably reveal the presence of many markets with one or more years of very large profits, but few instances of very large single-year losses. The implication is that a key reason for the success of these systems is that their rules adhere to the critical, albeit clichéd principle of letting profits run and cutting losses short.
10. A market should not be avoided because its volatility increases sharply. In fact, the most volatile markets are often the most profitable.
11. Isolating negative results for a system that performs well on balance can provide valuable clues as to how the system can be improved.
12. A frequently overlooked fact is that trading results may often reflect more information about the market than the system. For example, in Figure 19.3, the fact that a trend-following system that was short in mid-January 2015 would have witnessed the transformation of a large open profit into a large loss before the system provided a liquidation or reversal signal would not necessarily reflect inadequate risk control. Virtually any trend-following system would have experienced the same fate.

    This example illustrates how the value of a system cannot be judged in a vacuum. In some cases, poor performance may reflect nothing more than the fact that market conditions would have resulted in poor results for the vast majority of systems. Similarly, favorable results may also reflect the conditions of the market rather than any degree of superiority in the tested

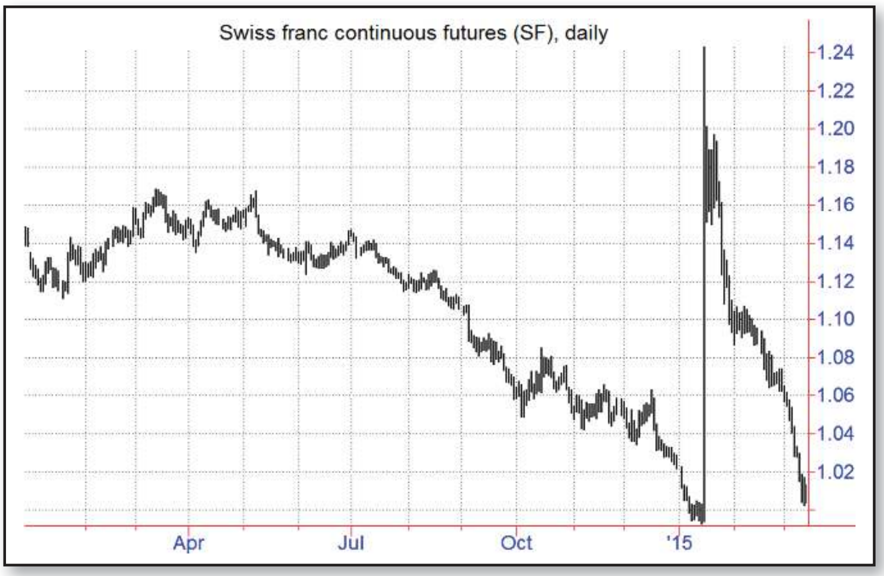

FIGURE 19.3 Trading Results Reflect Market, Not System: Short Swiss Franc Continuous Futures

Chart created using TradeStation. ©TradeStation Technologies, Inc. All rights reserved.

system. These considerations suggest that a meaningful assessment of a new system's performance should include a comparison to a benchmark (e.g., the corresponding performance of standard systems, such as a crossover moving average or a simple breakout, during the same period for the same markets).

13. Use continuous futures prices for testing systems.
14. Use only a small portion of the database (i.e., some markets for only a segment of the full time period) for developing and debugging a system.
15. Use charts with superimposed signal annotations as an aid to debugging systems.
16. In checking the accuracy and completeness of the signals generated by a system, make changes dictated by deviations from the intended operation of the system (due to oversights related to the full implications of the rules employed or unforeseen situations) with complete disregard for whether such changes increase or decrease profits in the sample tests.

CHAPTER 20

How to Evaluate Past Performance*

Why Return Alone Is Meaningless

You are looking for a London hotel room on the Internet. You find the same hotel room at two different sites (both including taxes) at two different prices:

- Site A: 300
- Site B: 250

Which is the better deal? The answer may seem obvious, but it's not. On one occasion, when I posed this question to a conference audience, one attendee shouted the response, "It depends whether they both include breakfast." "That would have to be a very expensive breakfast," I answered. But at least he had the right idea. The question I posed contained incomplete information. I didn't specify what currency the prices were quoted in. What if the 300 price was in dollars and the 250 price was in pounds (let's say when the pound was at $1.40)? Changes everything, doesn't it?

"Well," you are probably thinking, "no rational person will ignore the currency denomination in comparing two prices, so what's the point?" The point is that investors make this type of error all the time when selecting investments by focusing only on returns. Comparing returns without risk is as meaningless as comparing international hotel prices without the currency denomination. Risk is the denomination of return.

---

*This chapter is adapted from Jack D. Schwager, *Market Sense and Nonsense: How the Markets Really Work (and How They Don't)* (Hoboken, NJ: John Wiley & Sons, 2012).

TABLE 20.1 A Comparison of Two Managers

	Return	Risk (Standard Deviation)	Return/Risk Ratio
Manager A	10%	5	2:1
Manager B	25%	25	1:1

Consider the two managers in Table 20.1. Assuming the two managers are considered qualitatively equivalent, which is the better-performing manager?[1] Many investors would opt for Manager B, reasoning, "I am willing to accept the higher risk to get the higher return potential." But is this reasoning rational? In Table 20.2 we add a third investment alternative—leveraging an investment with Manager A at 300 percent.[2] The leveraged investment with Manager A now has both a higher return and lower risk than Manager B. So even risk-seeking investors should prefer Manager A, using a leverage factor that raises return to the desired level.

One can picture risk as a hole—the deeper the hole, the greater the risk—and return as a pile of sand. Leverage is the shovel that, if desired, allows transferring some of the sand from the risk hole to the return pile, thereby increasing return in exchange for accepting greater risk—a trade-off that may be preferred if the risk level is lower than desired. Continuing the analogy, by using negative leverage (i.e., holding more cash), it is also possible to transfer sand from the return pile to the risk hole, thereby reducing risk in exchange for accepting lower return. In this sense, risk and return are entirely interchangeable through leverage (that is, through varying exposure).

TABLE 20.2 A Comparison of Two Managers Revisited

	Return	Risk (Standard Deviation)	Return/Risk Ratio
Manager A	10%	5	2:1
Manager B	25%	25	1:1
Manager A 3×	30%	15	2:1

[1] Although this chapter is written from the perspective of an investor comparing investments with two different managers, exactly analogous comments would apply to a trader comparing two different systems or two different trading strategies.

[2] For strategies that use margin (e.g., futures, foreign exchange, options), managers need only a small percentage of the nominal investment to meet margin requirements. In these instances, investors can often use notional funding—that is, funding an account with a smaller amount of cash than the nominal level. For example, an investor might notionally fund an account with $300,000 cash to be traded as a $900,000 investment, implicitly leveraging the cash investment 300 percent vis-à-vis an investment that is not notionally funded. Technically speaking, although notional funding increases the exposure per dollar invested, it does not actually imply leverage, since there is no borrowing involved. Our example assumes notional funding. Nevertheless, in the ensuing discussion, we use the term leverage to indicate increased exposure (even if there is no borrowing involved). For strategies that must be fully funded, the leveraged portion of returns would have to be reduced by borrowing costs.

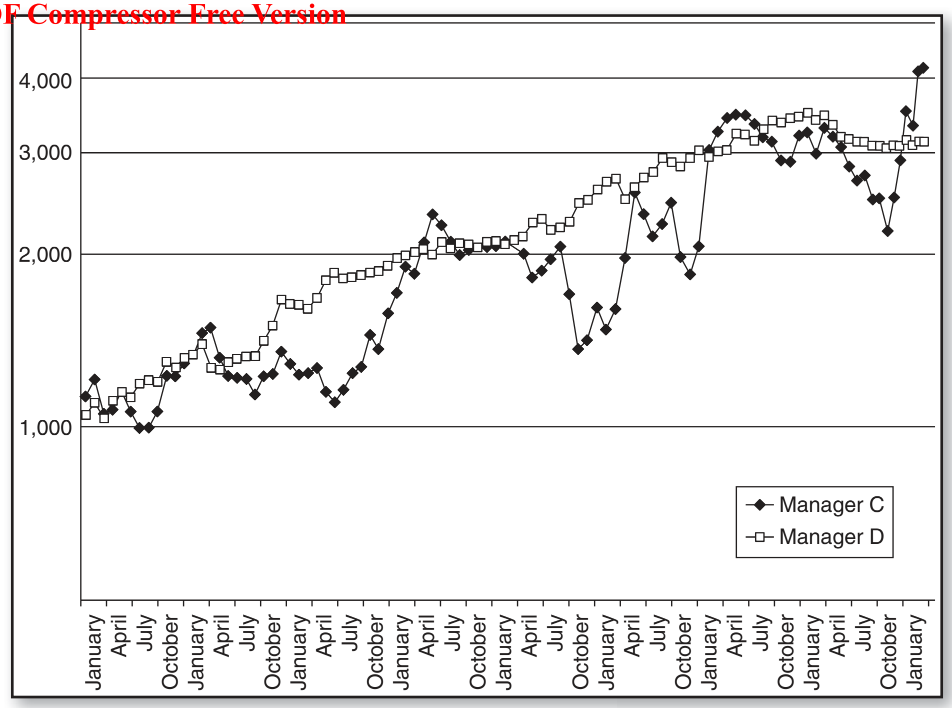

FIGURE 20.1 Two Paths to Return

As a practical example to illustrate this concept, in Figure 20.1 we compare two actual managers. Assuming we consider past performance indicative of potential future performance—at least in a relative sense—which manager provides a better investment? It would appear the answer is indeterminate: Manager C clearly achieves a superior return, but Manager D displays considerably lower risk, as evidenced by much smaller equity drawdowns throughout the track record. The seeming inability to determine which manager exhibits better performance is true only in a superficial sense, however. In Figure 20.2, we again compare Managers C and D, but this time we assume the exposure to Manager D is doubled.[3] Now it is clear that Manager D is superior in terms of both return and risk, achieving a significantly higher ending net asset value (NAV) and still doing so with visibly lower equity drawdowns (despite the doubling of exposure). Even though Manager C ended up with a higher return in Figure 20.1, investors could have achieved an even higher return with a 2× investment in Manager D while still maintaining less risk. The lesson is that return is a faulty gauge; it is the return/risk ratio that matters.

---

[3] Managers C and D are commodity trading advisors (CTAs) who trade futures, so increased exposure could have been achieved through notional funding (i.e., without leverage through borrowing). The returns depicted in Figure 20.1 were adjusted to remove interest income, so that doubling exposure (whether through notional funding or through borrowed leverage) would multiply all the returns by a near-exact factor of 2.0. (If returns included interest income, then doubling the exposure would not fully double the returns because there would be no interest income on the additional exposure.)

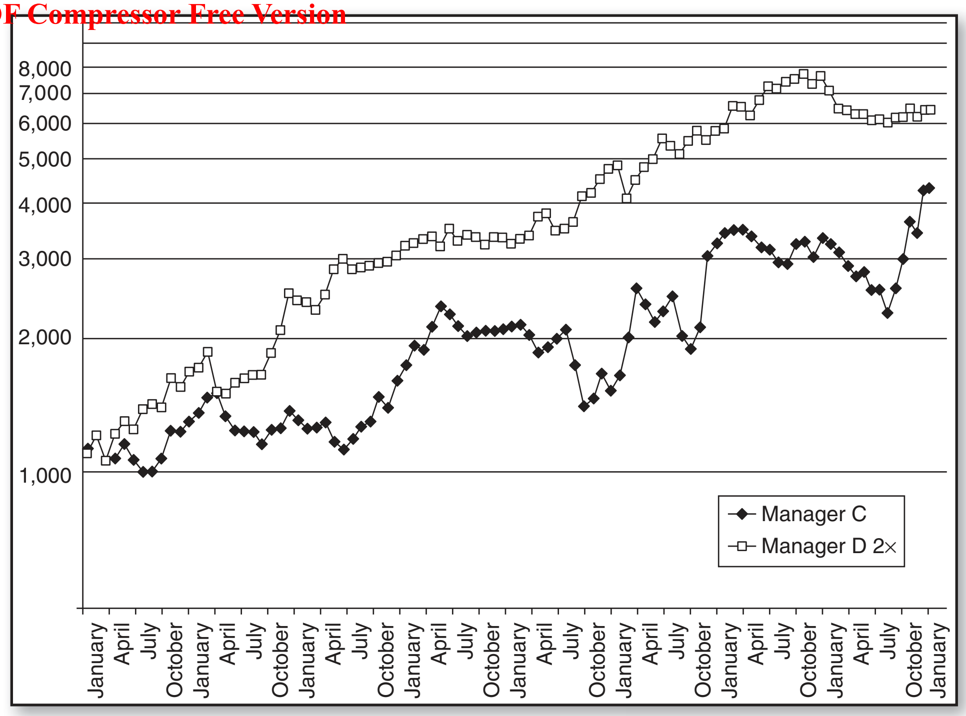

FIGURE 20.2 Doubling the Exposure of the Lower-Risk Manager

What if leverage is not available as a tool? For example, what if investors have a choice between Managers C and D in Figure 20.1 but there are practical impediments to increasing the exposure of Manager D? Now return and risk are inextricably bundled, and investors must choose between the higher-return/higher-risk profile of Manager C and the lower-return/lower-risk profile of Manager D. It might seem that risk-tolerant investors would always be better off with Manager C. Such investors might say, "I don't care if Manager C is riskier, as long as the end return is higher." The flaw in this premise is that investors who start with Manager C at the wrong time—and that is easy to do—may actually experience significant losses rather than gains, even if they maintain the investment, and especially if they don't. The more volatile the path of returns, the more likely investors will abandon the investment during one of the equity plunges and, as a result, never realize the higher return. After all, investors in real time do not know the investment will eventually recover. Thus, even though Manager C ends up ahead of Manager D, many investors will never survive the ride to see the eventual successful outcome (and even those who do may have initiated their investment on an upside excursion, reducing or even eliminating their net return). The greater the volatility, the larger the percentage of investors who will close out their investments at a loss.

Clearly, there is a need to use risk-adjusted returns rather than returns alone to make valid performance comparisons. In the next section we consider some alternative risk-adjusted return measures.

Risk-Adjusted Return Measures

Sharpe Ratio

The Sharpe ratio is the most widely used risk-adjusted return measure. The Sharpe ratio is defined as the average *excess return* divided by the standard deviation. Excess return is the return above the risk-free return (e.g., the Treasury bill rate). For example, if the average return is 8 percent per year and the T-bill rate is 3 percent, the excess return would be 5 percent. (It should be noted that during certain periods, such as the years following the 2008 financial crisis, zero, or near-zero, interest rates can effectively eliminate the expectation of a meaningful "risk-free" return. For reference, the average three-month T-bill rate from 2009 through 2015 was only 0.08 percent. In contrast, from 2002 to 2008 the average three-month T-bill rate was 2.58 percent, and during 1995–2001 it was 5.03 percent.) The standard deviation is a measure of the variability of return. In essence, the Sharpe ratio is the average excess return normalized by the volatility of returns:

$$SR = \frac{AR - RF}{SD}$$

where SR = Sharpe ratio
AR = average return (used as proxy for expected return)
RF = risk-free interest rate (e.g., Treasury bill return)
SD = standard deviation

The standard deviation is calculated as follows:

$$SD = \sqrt{\frac{\sum_I^N (X_i - \overline{X})^2}{N-1}}$$

where $\overline{X}$ = mean
X_i = individual returns
N = number of returns

Assuming monthly data is used to calculate the Sharpe ratio, as is most common, the Sharpe ratio would be annualized by multiplying by the square root of 12. Note that the return is an arithmetic average return, not the compounded return.

There are two basic problems with the Sharpe ratio:

1. **The return measure is based on average rather than compounded return.** The return an investor realizes is the compounded return, not the average return. The more volatile the return series, the more the average return will deviate from the actual (i.e., compounded) return. For example, a two-year period with a 50 percent gain in one year and a 50 percent loss in the other would represent a zero percent average return, but the investor would actually realize a 25 percent loss (150% × 50% = 75%). The average annual compounded return of –13.4 percent, however, would reflect the reality (86.6% × 86.6% = 75%).

2. **The Sharpe ratio does not distinguish between upside and downside volatility.** The risk measure inherent in the Sharpe ratio—the standard deviation—does not reflect the way most investors perceive risk. Investors care about loss, not volatility. They are averse to downside volatility, but actually like upside volatility. I have yet to meet any investors who complained because their managers made too much money in a month. The standard deviation, and by inference the Sharpe ratio, however, makes no distinction between upside and downside volatility. This characteristic of the Sharpe ratio can result in rankings that would contradict most investors' perceptions and preferences.[4]

Figure 20.3 compares two hypothetical managers that have identical returns over the period depicted, but very different return profiles. Which manager appears riskier? Decide on an answer before reading on.

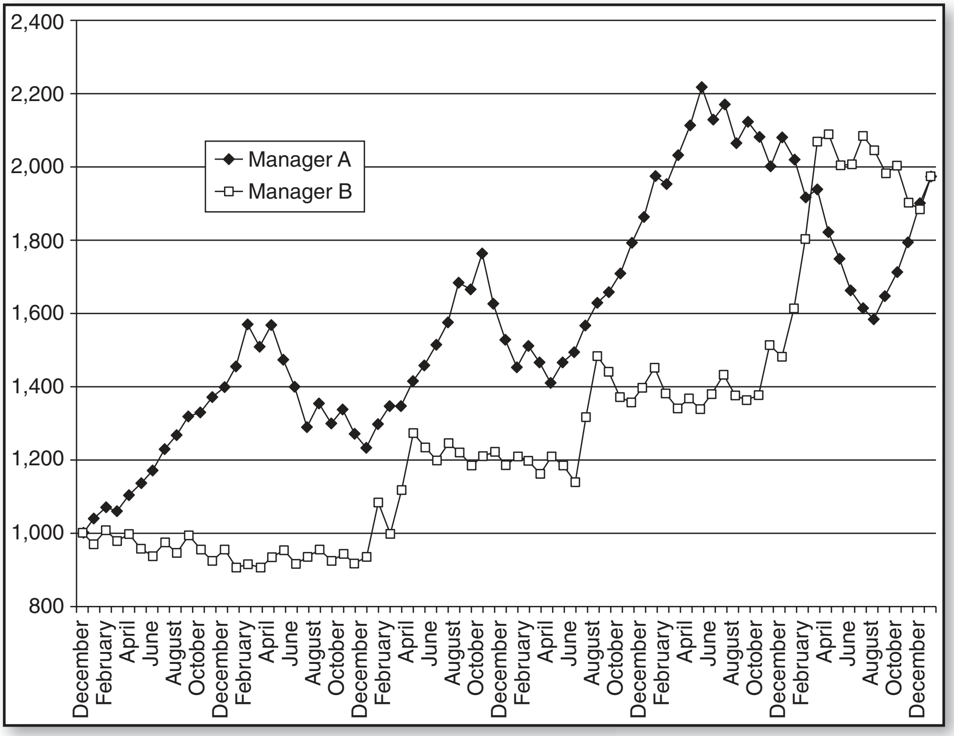

FIGURE 20.3 Which Manager Is Riskier?

[4]To be fair, in some cases, high upside volatility can be indicative of a greater potential for downside volatility, and in these instances the Sharpe ratio will be an appropriate measure. The Sharpe ratio, however, will be particularly misleading in evaluating strategies that are designed to achieve sporadic large gains while strictly controlling downside risk (that is, "right-skewed" strategies).

Most likely you chose Manager A as being riskier. Manager A has three drawdown episodes in excess of 20 percent, with the largest being 28 percent. In contrast, Manager B's worst peak-to-valley decline is a rather moderate 11 percent. Yet the standard deviation—the risk component of the Sharpe ratio—is 30 percent higher for Manager B. As a result, even though both Managers A and B have equal cumulative returns and Manager A has much larger equity retracements, Manager A also has a significantly higher Sharpe ratio: 0.71 versus 0.58 (assuming a 2 percent risk-free rate). Why does this occur? Because Manager B has a number of very large gain months, and it is these months that strongly push up Manager B's standard deviation, thereby reducing the Sharpe ratio. Although most investors would clearly prefer the return profile of Manager B, the Sharpe ratio decisively indicates the reverse ranking.

The potential for a mismatch between Sharpe ratio rankings and investor preferences has led to the creation of other return/risk measures that seek to address the flaws of the Sharpe ratio. Before we review some of these alternative measures, we first consider the question: What are the implications of a negative Sharpe ratio?

Although it is commonplace to see negative Sharpe ratios reported for managers whose returns are less than the risk-free return, negative Sharpe ratios are absolutely meaningless. When the Sharpe ratio is positive, greater volatility (as measured by the standard deviation), a negative characteristic, will reduce the Sharpe ratio, as it logically should. When the Sharpe ratio is negative, however, greater volatility will actually increase its value—that is, the division of a negative return by a larger number will make it less negative. Comparisons involving negative Sharpe ratios can lead to absurd results. Table 20.3 provides an example. Manager B has a negative excess return twice the size of Manager A's (–10 percent versus –5 percent) and four times the volatility of Manager A. Even though Manager B is much worse than Manager A in terms of both return and volatility, Manager B has a higher (less negative) Sharpe ratio. This preposterous result is a direct consequence of higher volatility resulting in higher (less negative) Sharpe ratios when the Sharpe ratio is in negative territory. What should be done with negative Sharpe ratios? Ignore them.[5] They are always worthless and frequently misleading.

Sortino Ratio

The Sortino ratio addresses both of the Sharpe ratio's previously cited problems. First, it uses the compounded return, which is representative of the actual realized return over any period of time,

TABLE 20.3 A Comparison of Two Managers with Negative Sharpe Ratios

	Average Annual Return	Risk-Free Return	Excess Return	Annualized Standard Deviation	Sharpe Ratio
Manager A	–3%	2%	–5%	5	–1.0
Manager B	–8%	2%	–10%	20	–0.5

[5] What if some value must be used, as in an application such as ranking a list of managers based on the ratio? In this case, a dual rank criterion makes much more sense: ranking managers based on the Sharpe ratio when excess returns are positive and on excess returns when Sharpe ratios are negative.

instead of the arithmetic return. Second, and most important, the Sortino ratio focuses on defining risk in terms of downside deviation, considering only deviations below a specified minimum acceptable return (MAR) instead of a standard deviation (used in the Sharpe ratio), which includes all deviations, upside as well as downside. Specifically, the Sortino ratio is defined as the compounded return in excess of the MAR divided by the downside deviation, as follows:

$$SR = \frac{ACR - MAR}{DD}$$

where SR = Sortino ratio
ACR = annual compounded return
MAR = minimum acceptable return (e.g., zero, risk-free, average)
DD = downside deviation

where DD is defined as:

$$DD = \sqrt{\frac{\sum_i^N (\min(X_i - MAR, 0))^2}{N}}$$

where X_i = individual returns
MAR = minimum acceptable return (e.g., zero, risk-free, average)
N = number of data values

For example, if we define $MAR = 0$, then DD calculations will include only deviations for months with negative returns (the other months will equal zero).

The MAR in the Sortino ratio can be set to any level, but one of the following three definitions is normally used for the MAR:

1. **Zero.** Deviations are calculated for all negative returns.
2. **Risk-free return.** Deviations are calculated for all returns below the risk-free return.
3. **Average return.** Deviations are calculated for all returns below the average of the series being analyzed. This formulation is closest to the standard deviation, but considers deviations for only the lower half of returns.

Frequently, the fact that a manager has a higher Sortino ratio than Sharpe ratio is cited as evidence that returns are positively skewed—that is, there is a tendency for larger deviations on the upside than on the downside. This type of comparison is incorrect. The Sortino and Sharpe ratios cannot be compared, and as formulated, the Sortino ratio will invariably be higher, even for managers whose worst losses tend to be larger than their best gains. The reason for the upward bias in the Sortino ratio is that it calculates deviations for only a portion of returns—those returns below the MAR—but uses a divisor based on the number of *all* returns to calculate the downside deviation. Because it distinguishes between upside and downside deviations, the Sortino ratio probably comes closer to reflecting investor preferences than does the Sharpe ratio and, in this sense, may be a better tool for comparing managers. But the Sortino ratio should be compared only with other Sortino ratios and never with Sharpe ratios.

Symmetric Downside-Risk Sharpe Ratio

The symmetric downside-risk (SDR) Sharpe ratio, which was introduced by William T. Ziemba,[6] is similar in intent and construction to the Sortino ratio, but makes a critical adjustment to remove the inherent upward bias in the Sortino ratio vis-à-vis the Sharpe ratio. The SDR Sharpe ratio is defined as the compounded return minus the risk-free return divided by the downside deviation. The downside deviation is calculated similarly to the downside deviation in the Sortino ratio with one critical exception: a multiplier of 2.0 is used to compensate for the fact that only returns below a specified benchmark contribute to the deviation calculation.[7] The benchmark used for calculating the downside deviation can be set to any level, but the same three choices listed for the MAR in the Sortino ratio would apply here as well: zero, risk-free return, and average return. (In his article, Ziemba uses zero as the benchmark value.) Unlike the Sortino ratio, the SDR Sharpe ratio (with the benchmark set to the average) can be directly compared with the Sharpe ratio.[8]

$$SDRSR = \frac{ACR - RF}{\sqrt{2} \times DD}$$

where $SDRSR$ = symmetric downside-risk Sharpe ratio
ACR = annual compounded return
RF = risk-free interest rate (e.g., T-bill return)
DD = downside deviation

[6] William T. Ziemba, "The Symmetric Downside-Risk Sharpe Ratio," *Journal of Portfolio Management* (Fall 2005): 108–121.

[7] Ziemba used the term *benchmark* instead of MAR in defining downside deviation. If the median were used as the benchmark, only half the returns would be used to calculate the downside deviation, and a multiplier of 2.0 would then provide an exact compensating adjustment. For other choices for the benchmark (e.g., zero, risk-free return, average), the number of points below the benchmark would not necessarily be exactly half, and a multiplier of 2.0 would provide an approximate adjustment.

[8] To be perfectly precise, there would be a tendency for the SDR Sharpe ratio to be slightly lower for a symmetric distribution of returns because the SDR Sharpe ratio uses the compounded return rather than the arithmetic return used in the Sharpe ratio, and the arithmetic return will always be equal to or higher than the compounded return. If, however, zero or the risk-free return is used as the benchmark in the downside deviation calculation, assuming the manager's average return is greater than the risk-free return, there would be a tendency for the SDR Sharpe ratio to be higher than the Sharpe ratio for a symmetric distribution of returns for two reasons:

1. There will be fewer than half the returns below the benchmark, so the multiplication by 2.0 will not fully compensate.
2. Downside deviations from the risk-free return (and especially zero) would be smaller than deviations from the average.

These two factors would cause the downside deviation to be smaller than the standard deviation, implying a higher SDR Sharpe ratio than Sharpe ratio.

where DD is defined as:

$$DD = \sqrt{\frac{\sum_i^N (\min(X_i - \overline{X}, 0))^2}{N - 1}}$$

where X_i = individual returns

$\overline{X}$ = benchmark return (e.g., mean, zero, risk-free)

Since the SDR Sharpe ratio includes only the downside deviation, multiplying by the square root of 2 (a consequence of doubling the squared deviations) is equivalent to assuming the upside deviation is equal (i.e., symmetric) to the downside deviation. This proxy replacement of the upside deviation is what makes it possible to compare SDR Sharpe ratio values with Sharpe ratio values.

The SDR Sharpe ratio (with any of the standard choices for a benchmark value) is preferable to the Sharpe ratio because it accounts for the very significant difference between the risk implications of downside deviations versus upside deviations as viewed from the perspective of the investor. The SDR Sharpe ratio is also preferable to the Sortino ratio because it is an almost identical calculation,[9] but with the important advantage of being directly comparable with the widely used Sharpe ratio. Also, by comparing a manager's SDR Sharpe ratio versus the Sharpe ratio, an investor can get a sense of whether the manager's returns are positively or negatively skewed.

Gain-to-Pain Ratio

The gain-to-pain ratio (GPR) is the sum of all monthly returns divided by the absolute value of the sum of all monthly losses.[10] This performance measure indicates the ratio of cumulative net gain to the cumulative loss realized to achieve that gain. For example, a GPR of 1.0 would imply that, on average, an investor has to experience an amount of monthly losses equal to the net amount gained. The GPR penalizes all losses in proportion to their size, and upside volatility is beneficial since it impacts only the return portion of the ratio.

[9] Besides the essential introduction of the 2.0 multiplier term, which allows unbiased comparisons between the SDR Sharpe ratio and the Sharpe ratio, the only difference between the SDR Sharpe ratio and the Sortino ratio is that it subtracts the risk-free return from the compounded return instead of the MAR (which may or may not be the risk-free return).

[10] The gain-to-pain ratio (GPR) is a performance statistic I have been using for many years. I am not aware of any prior use of this statistic, although the term is sometimes used as a generic reference for return/risk measures or a return/drawdown measure. The GPR is similar to the profit factor, which is a commonly used statistic in evaluating trading systems. The profit factor is defined as the sum of all profitable trades divided by the absolute value of the sum of all losing trades. The profit factor is applied to trades, whereas the GPR is applied to interval (e.g., monthly) returns. Algebraically, it can easily be shown that if the profit factor calculation were applied to monthly returns, the profit factor would equal GPR + 1 and would provide the same performance ordering as the GPR. For quantitatively oriented readers familiar with the omega function, note that the omega function evaluated at zero is also equal to GPR + 1.

$$GPR = \frac{\sum_{i=1}^{N} X_i}{\left| \sum_{i}^{N} \min(X_i, 0) \right|}$$

where X_i = individual returns

A key difference between the GPR and measures such as the Sharpe ratio, the SDR Sharpe ratio, and the Sortino ratio is that the GPR will be indifferent between five 2 percent losses and one 10 percent loss, whereas the other ratios discussed so far will be impacted far more by the single larger loss. This difference results because the standard deviation and downside deviation calculations used for the other ratios involve squaring the deviation between the reference return level (e.g., average, zero, risk-free) and the loss. For example, if the reference return is zero percent, the squared deviation for one 10 percent loss would be five times greater than the squared deviation for five 2 percent losses ($10^2 = 100$; $5 \times 2^2 = 20$). In the GPR calculation, by contrast, both cases will add 10 percent to the denominator. If an investor is indifferent as to whether a given magnitude of loss is experienced over multiple months or in a single month, then the GPR would be a more appropriate measure than the SDR Sharpe ratio and Sortino ratio. However, an investor who considers a single larger loss worse than multiple losses totaling the same amount would have the opposite preference.

Although the GPR would typically be applied to monthly data, it can also be calculated for other time intervals. If daily data are available, the GPR can provide a statistically very significant measure because of the large amount of sample data. The longer the time frame, the higher the GPR because many of the losses visible on a shorter time interval will be smoothed out over a longer period. In my experience, *on average*, daily GPR values tend to be about one-sixth as large as the monthly GPR for the same manager, although the ratio between daily and monthly GPR values can range widely. For monthly data, roughly speaking, GPRs greater than 1.0 are good and those above 1.5 are very good. For daily data, the corresponding numbers would be approximately 0.17 and 0.25.

One advantage of the GPR over the other ratios is that rankings remain consistent even for negative returns—that is, a smaller negative GPR is always better than a larger negative GPR (a relationship that is not necessarily true for the other ratios). A GPR of zero means that the sum of all wins is equal to the sum of all losses. The theoretical minimum GPR value is –1.0 and would occur if there were no winning months. The closer the GPR is to –1.0, the smaller the ratio of the sum of all wins to the sum of all losses.[11]

Tail Ratio

An important question for the investor is whether a manager's extreme returns tend to be larger on the upside or the downside. Managers with frequent small gains and occasional large losses (negatively skewed managers) are more risky and less desirable than managers with frequent small

[11] The ratio of the sum of wins to the sum of losses is equal to GPR + 1. So, for example, a GPR of −0.25 would imply that the ratio of the sum of wins to the sum of losses is 0.75.

losses and occasional large gains (positively skewed managers). Although there is a statistic that measures skewness—the degree to which a return distribution has longer tails (extreme events) on the right (positive) or left (negative) side than the symmetric normal distribution—it is difficult to attach intuitive meaning to specific values (beyond the value of the sign).

The tail ratio measures the tendency for extreme returns to be skewed to the positive or negative side in a statistic whose value is intuitively clear.

$$TR = \frac{\dfrac{\sum_{p=0}^{p=T} X_p}{N_{p<T}}}{\left| \dfrac{\sum_{p=100-T}^{p=100} X_p}{N_{p>100-T}} \right|}$$

where X_p = return at percentile p

T = threshold percentile to calculate numerator of tail ratio (Implicit assumption: Lower percentile rankings represent higher return. For example, the top 10% of returns would be all returns less than T, where $T = 10$.)

$N_{p<T}$ = number of returns below percentile

$N_{p>100-T}$ = number of returns above percentile $100-T$

The tail ratio requires one parameter input: the upper and lower percentile threshold used to calculate the statistic. If the threshold is set to 10, for example, the tail ratio would be equal to the average of all returns in the top decile of returns divided by the absolute value of the average of all returns in the bottom decile of returns. (*Note:* If the average of bottom decile returns is positive, the tail ratio would have no meaning and cannot be calculated.) If returns were normally distributed, the tail ratio would equal 1.0. A ratio significantly less than 1.0 would indicate a tendency for the largest losses to be of greater magnitude than the largest gains, while a ratio significantly greater than 1.0 would indicate the reverse tendency. For example, if the tail ratio was equal to 0.5, it would imply that the magnitude of the average loss in the bottom decile was twice as large as the average gain in the top decile—a reading indicative of a potentially very risky manager.

MAR and Calmar Ratios

The MAR ratio is the annualized compounded return divided by the maximum drawdown.

$$MAR = \frac{ACR}{1 - \min\left(\dfrac{NAV_j}{NAV_i}\right)}$$

where ACR = annual compounded return (expressed in decimal form)

NAV = net asset value

$j > i$

The Calmar ratio is exactly the same except the calculation is specifically restricted to the past three years of data. Although these ratios are useful in that they are based on a past worst-case situation, the fact that the risk measure divisor is based on only a single event impedes their statistical significance. Also, if applied over entire track records, the MAR will be strongly biased against managers with longer records, because the longer the record, the greater the potential maximum drawdown. (This bias does not exist in the Calmar ratio because, by definition, it is based on only the past three years of data.) Manager comparisons should be limited to common time periods, a restriction that is especially critical when using the MAR ratio.

Return Retracement Ratio

The return retracement ratio (RRR) is similar to the MAR and Calmar ratios in that it is a measure of the average annual compounded return divided by a retracement measure. The key difference, however, is that instead of being based on a single retracement (the maximum retracement), the RRR divides return by the average maximum retracement (AMR), which is based on a maximum retracement calculation for each month. The maximum retracement for each month is equal to the greater of the following two numbers:

1. The largest possible cumulative loss that could have been experienced by any existing investor in that month (the percentage decline from the prior peak NAV to the current month-end NAV).
2. The largest loss that could have been experienced by any new investor starting at the end of that month (the percentage decline from the current month-end NAV to the subsequent lowest NAV).

$$RRR = \frac{ACR - RF}{AMR}$$

where ACR = annual compounded return
RF = risk-free return
AMR = average maximum retracement = MR_i / N

where N = number of months
MR_i = max($MRPNH_i$, $MRSNL_i$)

where $MRPNH_i$ is the maximum retracement from prior NAV high, and is defined as:

$$MRPNH_i = (PNH_i - NAV_i) / PNH_i$$

where PNH_i = prior NAV high (prior to month i)
NAV_i = NAV at end of month i

$MRSNL_i$ is the maximum retracement to a subsequent NAV low, and is defined as:

$$MRSNL_i = (NAV_i - SNL_i) / NAV_i$$

where SNL_i is the subsequent NAV low (subsequent to month i).

The reason for using both metrics to determine a maximum retracement for each month is that each of the two conditions would be biased to show small retracement levels during a segment of the track record. The first condition would invariably show small retracements for the early months in the track record because there would not have been an opportunity for any large retracements to develop. Similarly, the second condition would inevitably show small retracements during the latter months of the track record for analogous reasons. By using the maximum of both conditions, we assure a true worst-case number for each month. The average maximum retracement is the average of all these monthly maximum retracements. The return retracement ratio is statistically far more meaningful than the MAR and Calmar ratios because it is based on multiple data points (one for each month) as opposed to a single statistic (the maximum drawdown in the entire record).

Comparing the Risk-Adjusted Return Performance Measures

Table 20.4 compares Managers A and B shown in Figure 20.3 in terms of each of the risk-adjusted return performance measures we discussed. Interestingly, the Sharpe ratio, which is by far the most widely used return/risk measure, leads to exactly the opposite conclusion indicated by all the other measures. Whereas the Sharpe ratio implies that Manager A is significantly superior in return/risk terms, all the other performance measures rank Manager B higher—many by wide margins. Recall that both Managers A and B had identical cumulative returns, so the only difference between the two was the riskiness implied by their return paths. The Sharpe ratio, which uses the standard deviation as its risk metric, judged Manager B as being riskier because of higher volatility, as measured across all months. Most of Manager B's volatility, however, was on the upside—a

TABLE 20.4 **A Comparison of Risk-Adjusted Return Measures**

	Manager A	Manager B	B as Percent of A
Sharpe ratio	0.71	0.58	82%
Sortino ratio (zero)	1.27	1.44	113%
Sortino ratio (risk-free)	1.03	1.15	112%
Sortino ratio (average)	0.87	0.94	107%
SDR Sharpe ratio (zero)	0.75	0.85	113%
SDR Sharpe ratio (risk-free)	0.73	0.81	112%
SDR Sharpe ratio (average)	0.62	0.66	107%
Gain-to-pain ratio (GPR)	0.70	0.71	101%
Tail ratio (10%)	1.13	2.86	253%
Tail ratio (5%)	1.10	2.72	247%
MAR ratio	0.41	1.09	265%
Calmar ratio	0.33	1.70	515%
Return retracement ratio (RRR)	0.77	1.67	218%

characteristic most investors would consider an attribute, not a fault. Although Manager A had lower volatility overall, the downside volatility was significantly greater than Manager B's—a characteristic that is consistent with most investors' intuitive sense of greater risk. The Sharpe ratio does not distinguish between downside and upside volatility, while the other risk-adjusted return measures do.

Although all the risk-adjusted return measures besides the Sharpe ratio penalize only downside volatility, they do so in different ways that have different implications:

- **Sortino ratio and SDR Sharpe ratio.** These ratios penalize returns below a specified level (e.g., zero) with the weight assigned to downside deviations increasing more than proportionately as their magnitude increases. Thus, one larger downside deviation will reduce the ratio more than multiple smaller deviations that sum to the same amount. These ratios are unaffected by the order of losing months. Two widely separated losses of 10 percent will have the same effect as two consecutive 10 percent losses, even though the latter results in a larger equity retracement.

- **GPR.** The GPR penalizes downside deviations in direct proportion to their magnitude. In contrast to the Sortino and SDR Sharpe ratios, one large deviation will have exactly the same effect as multiple smaller deviations that sum to the same amount. This difference explains why Managers A and B are nearly equivalent based on the GPR, but Manager A is significantly worse based on the Sortino and SDR Sharpe ratios: Manager A has both larger and fewer losses, but the sum of the losses is nearly the same for both managers. The GPR is similar to the Sortino and SDR Sharpe ratios in terms of being indifferent to the order of losses; that is, it does not penalize for consecutive or proximate losses.

- **Tail ratio.** The tail ratio focuses specifically on the most extreme gains and losses. The tail ratio will be very effective in highlighting managers whose worst losses tend to be larger than their best gains. In terms of the tail ratio, Manager B, who achieves occasional very large gains but whose worst losses are only moderate, is dramatically better than Manager A, who exhibits the reverse pattern.

- **MAR and Calmar ratios.** In contrast to all the foregoing performance measures, these ratios are heavily influenced by the order of returns. A concentration of losses will have a much greater impact than the same losses dispersed throughout the track record. Both of these measures, however, focus on only the single worst equity drawdown. Therefore losses that occur outside the interim defined by the largest peak-to-valley equity drawdown will not have any impact on these ratios. Because the maximum drawdown for Manager A is much greater than for Manager B, these ratios show a dramatic difference between the two managers.

- **Return retracement ratio (RRR).** The RRR is the only return/risk measure that both penalizes *all* downside deviations and also penalizes consecutive or proximate losses. In contrast to the MAR and Calmar ratios, which reflect only those losses that define the maximum drawdown, the RRR calculation incorporates all losses.

Table 20.5 summarizes and compares the properties of the different risk-adjusted return measures.

TABLE 20.5 Properties of Risk-Adjusted Performance Measures

Property	Sharpe Ratio	SDR Sharpe Ratio	Sortino Ratio	GPR	Tail Ratio	MAR and Calmar	RRR
Is impacted by upside volatility	X						
Is impacted only by downside volatility		X	X	X	X	X	X
Reflects *all* downside volatility	X	X	X	X			X
Gives more than proportionate weight to large losses	X	X	X		X		
Is impacted by proximity of losses						X	X
Focuses on extreme returns only					X		
Rankings remain consistent for net negative returns				X	X		

Which Return/Risk Measure Is Best?

To some extent, the choice of which return/risk measures to use depends on the performance measure properties favored by the individual investor. The major advantages and disadvantages of these performance measures can be summarized as follows:

- **Sharpe ratio.** Although the Sharpe ratio is the most widely used risk-adjusted metric, it provides rankings that are least consistent with most people's intuitive sense of risk because it penalizes upside gains.
- **Sortino ratio.** This ratio corrects the main deficiency of the Sharpe ratio by focusing on downside risk instead of total volatility as the measure of risk. In addition, the Sortino ratio uses a compounded return, which matches actual return over the entire period, whereas the Sharpe ratio uses an arithmetic average return, which does not. One disadvantage of the Sortino ratio is that it is not directly comparable with the Sharpe ratio because its calculation is biased to delivering higher values.
- **SDR Sharpe ratio.** This ratio provides the same fix as the Sortino ratio, and it has the advantage of an additional adjustment that allows for direct comparisons of its values with Sharpe ratio values. Similar to the Sortino ratio, the SDR Sharpe ratio also uses the compounded return instead of the arithmetic average return. Since the SDR Sharpe ratio will provide nearly identical rankings as the Sortino ratio and has the advantage of allowing for comparisons with the Sharpe ratio for the same manager, it seems the better choice for any investor. Using both ratios would be redundant.
- **Gain-to-pain ratio (GPR).** Similar to the Sortino and SDR Sharpe ratios, the GPR penalizes a manager only for losses (zero percent is also a common choice for minimum acceptable return or benchmark in the Sortino and SDR Sharpe ratios). The GPR weights losses proportionately to their

magnitude, whereas the Sortino and SDR Sharpe ratios magnify the weight of larger losses. Investors who view one 10 percent monthly loss the same as five 2 percent losses might prefer the GPR, whereas investors who consider the single 10 percent monthly loss to be worse might prefer the SDR Sharpe ratio.

- **Tail ratio.** Since, by definition, the tail ratio considers only a small percentage of all returns (20 percent or less), it is not intended as a stand-alone risk-adjusted return measure. Its focus on extreme returns, however, makes it a very useful supplemental metric to one of the other measures.
- **MAR and Calmar ratios.** These ratios will penalize for losses that occur with sufficient proximity to be part of the same drawdown. The other ratios (with the exception of the RRR) are unaffected by the sequence of returns. The drawback of these ratios is that the risk is defined by only a single event (the maximum drawdown), impeding their statistical significance and representativeness.
- **Return retracement ratio (RRR).** This ratio is both based on downside deviations and impacted by proximate losses. Its big advantage vis-à-vis the MAR and Calmar ratios is that it reflects all retracements, with the risk number based on all monthly numbers, rather than just a single event and single statistic: the maximum drawdown. Although the MAR and Calmar ratios might still be consulted as supplemental measures reflecting a worst-case situation, the RRR is preferable as a return/drawdown ratio.

Visual Performance Evaluation

Many people will find that the performance charts in this section provide a better intuitive sense of relative performance (in both return and risk terms) than do performance statistics.

Net Asset Value (NAV) Charts

An NAV chart, such as was illustrated in Figure 20.3, provides an extremely useful way of evaluating a track record. The NAV chart depicts the compounded growth of \$1,000 over time. For example, an NAV of 2,000 implies that the original investment has doubled from its starting level as of the indicated time. The NAV chart can offer a good intuitive sense of past performance in terms of both return and risk. In fact, if an investor were to examine only a single performance gauge, the NAV chart would probably be the most informative.

The way we visually perceive conventionally scaled NAV charts that depict longer-term periods, however, may result in misleading inferences. Consider Figure 20.4, and answer the following three questions before reading on:

1. Was return higher in the first half of the track record or the second?
2. Was the manager riskier during the first half of the track record or the second?
3. Was the return/risk performance better during the first half of the track record or the second?

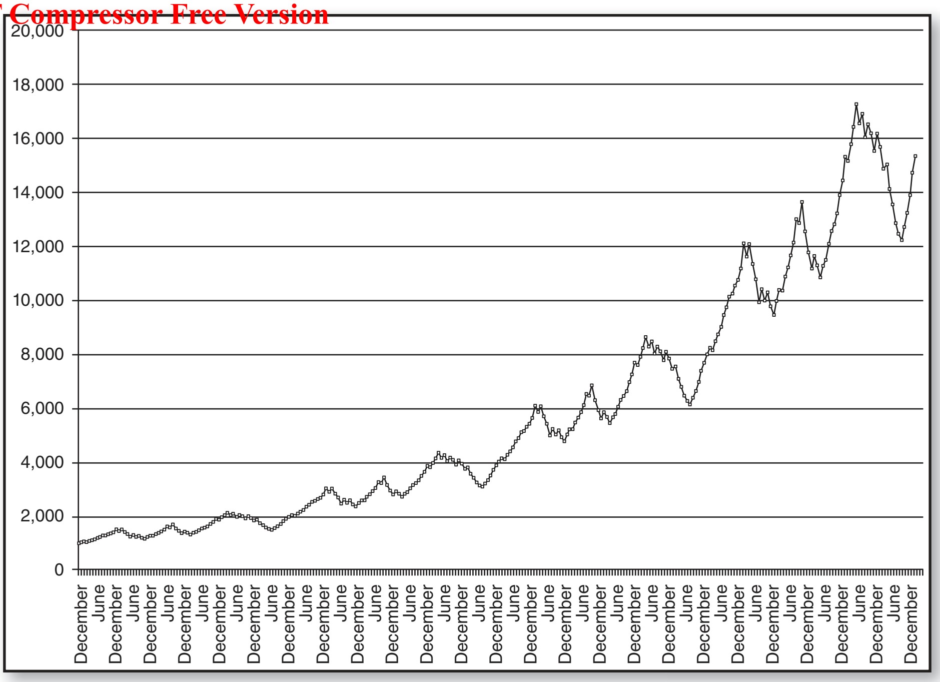

FIGURE 20.4 How Has Performance Changed over Time?

If you picked the first half as the answer to any of these three questions, you are wrong. If you picked the second half for any answer, you are also wrong. The two halves are exactly the same. In fact, all four quarters of the track record are the same. Figure 20.4 was created by copying the returns of Manager A in Figure 20.3 and pasting the sequence three times to the end to create an extended NAV that repeats the same return pattern, displaying it four times in all. Looking at Figure 20.4, however, it seems as if both the return and the volatility are increasing sharply over time. They are not. The illusion is an artifact of depicting NAV charts on a conventional arithmetic scale. On an arithmetic scale, an NAV decline of 1,000 when the NAV is at 16,000 looks the same as an NAV decline of 1,000 when the NAV is 2,000. The two declines, however, are radically different: a modest 6 percent decline in the first instance and a huge 50 percent drop in the second. The distortion on an arithmetic scale chart will get magnified when the NAV range is wide, which is frequently a serious problem for long-term charts.

The ideal way to depict an NAV chart is on a logarithmic scale. On a log scale chart, the increments for a fixed amount of movement (e.g., 1,000) become proportionately smaller as the level increases, and as a result, equal percentage price moves will appear as equal size moves on the vertical scale. Figure 20.5 depicts the same NAV as Figure 20.4, but on a log scale. The self-replicating nature of the chart is now evident as equal percentage changes now look identical wherever they appear. The moral is that a log scale is always the correct way to represent an NAV chart and is especially critical when there is a wide NAV range (more likely on long-term charts). A log scale was used for Figures 20.1 and 20.2 earlier in this chapter to allow for an accurate representation of relative volatility across time.

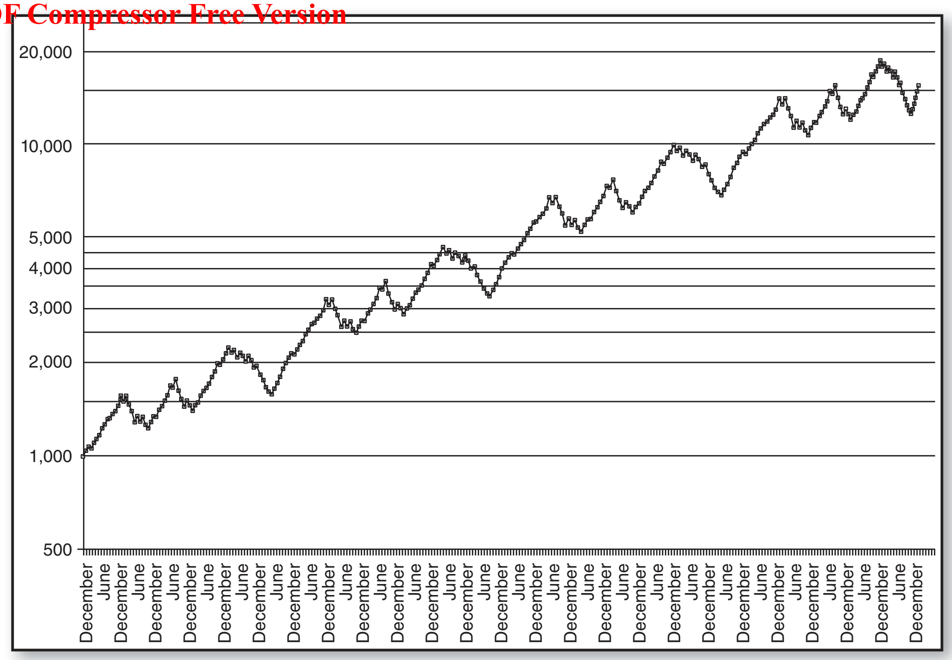

FIGURE 20.5 Log Scale: Equal Percentage Price Moves Appear Equal

Rolling Window Return Charts

The rolling window return chart shows the return for the specified time length ending in each month. For example, a 12-month rolling window return chart would show the 12-month return ending in each month (beginning with the 12th month of the track record). The rolling window return chart provides a clear visual summary of the results of investing with a manager for a specified length of time and answers such questions as: What would have been the range of outcomes with a manager for investments held for 12 months? 24 months? What was the worst loss for investments held for 12 months? 24 months?

For any December, the rolling 12-month return would be the same as the annual return. The important difference is that the rolling window return chart would show the analogous returns for all the other months as well. There is only a one-out-of-12 chance that December will be the worst 12-month return for the year. By showing all 12-month returns ending in any month, the rolling window chart will encompass worst-case events likely to be missed by annual returns and will provide a much more representative performance picture for one-year holding periods. The rolling window return chart can be calculated for other time intervals as well (e.g., 24 months, 36 months).

To illustrate the use of the rolling window return chart as a graphic analysis tool, we compare the two managers shown in Figure 20.6, who differ only moderately in terms of return (Manager E's annual compounded return is 1.3 percent higher), but differ widely in terms of the stability of returns. As shown in Figure 20.7, Manager E's 12-month returns range enormously from a severe

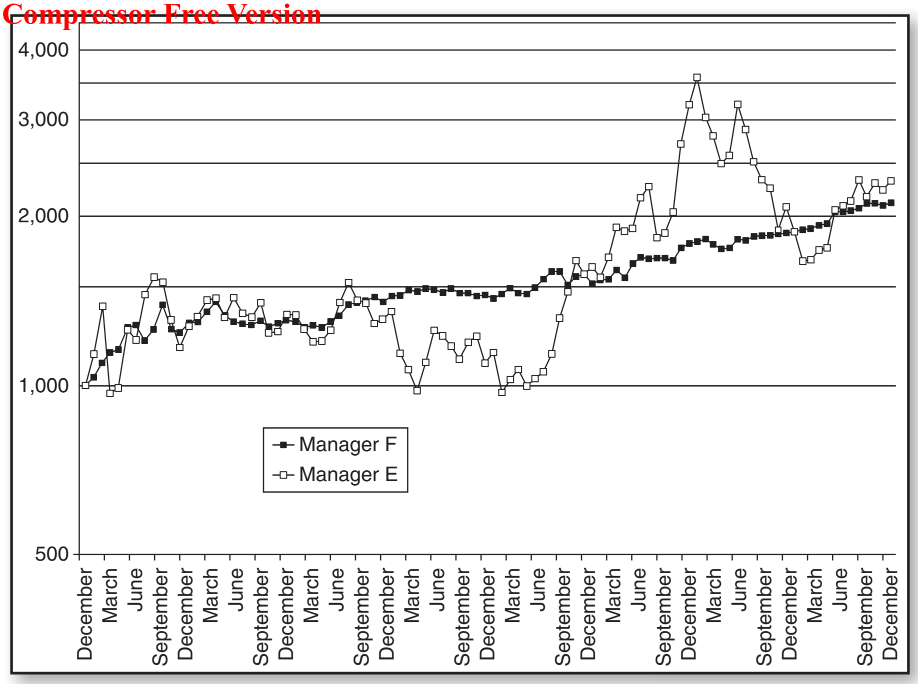

FIGURE 20.6 Small Difference in Return; Wide Difference in Stability of Return

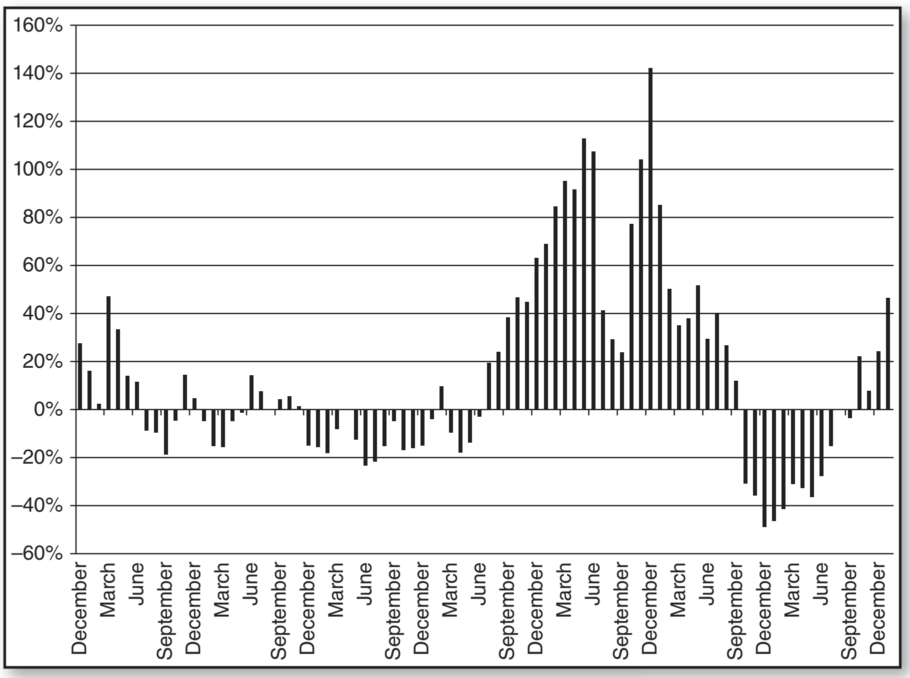

FIGURE 20.7 12-Month Rolling Return: Manager E

loss of 49 percent to a spectacular gain of 142 percent. In contrast, manager F's 12-month returns are contained in a far more moderate range of –10 percent to +29 percent (see Figure 20.8). Investors who were patient enough to stay with Manager F for at least 12 months would have experienced only a handful of investment initiation months that would have resulted in a net loss. Such patience, however, would not have provided any solace to investors with Manager E, who would have witnessed more than one-quarter of all 12-month holding periods resulting in net losses exceeding 15 percent, with several in excess of 40 percent. Even investors who committed to a 24-month holding period with Manager E would still have been subject to nearly one-fifth of all intervals with losses in excess of 15 percent (see Figure 20.9). In contrast, the worst-case outcome for investors with Manager F for a 24-month holding period would have been a positive return of 4 percent (see Figure 20.10).

Investors can use the rolling window return chart to assess the potential frequency and magnitude of worst-case outcomes as an aid in selecting investments consistent with their holding period tolerance for a losing investment. For example, an investor who is unwilling to maintain a losing investment for more than 12 months should avoid managers who have a meaningful percentage of negative 12-month returns, regardless of how favorable all the other performance statistics may be.

Rolling charts can also be used to depict other statistics besides return. For example, a rolling chart of annualized volatility (using daily data and a window of several months) can be used as a tool to monitor both managers and portfolios for early evidence of a possible increase in risk.

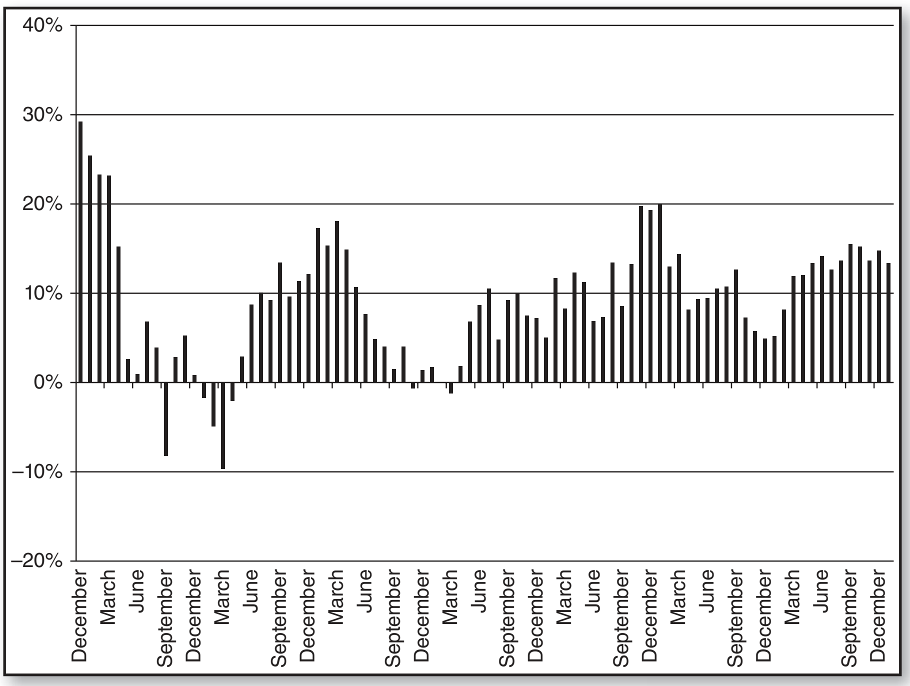

FIGURE 20.8 12-Month Rolling Return: Manager F

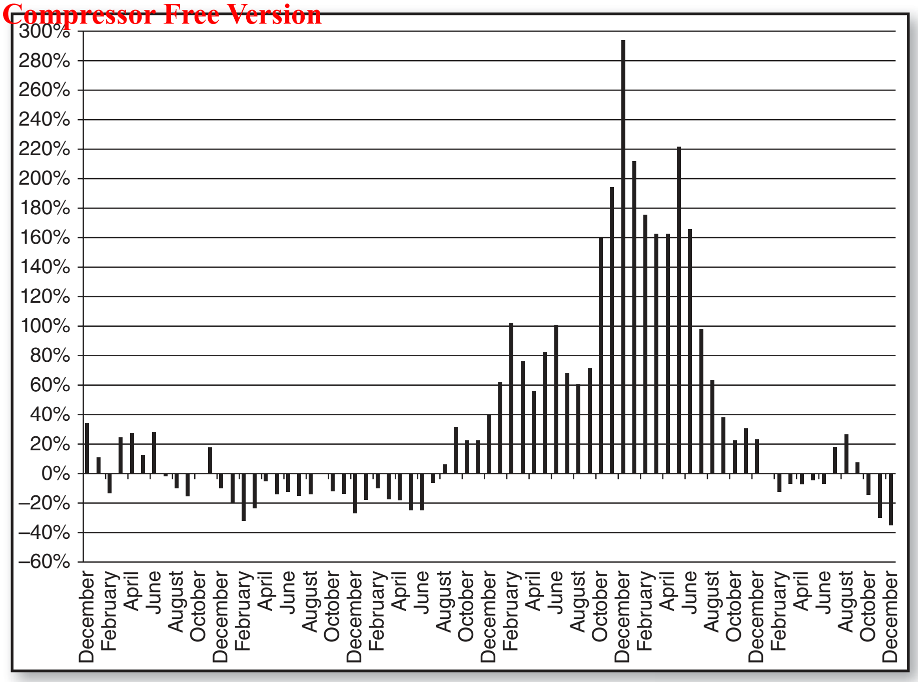

FIGURE 20.9 24-Month Rolling Return: Manager E

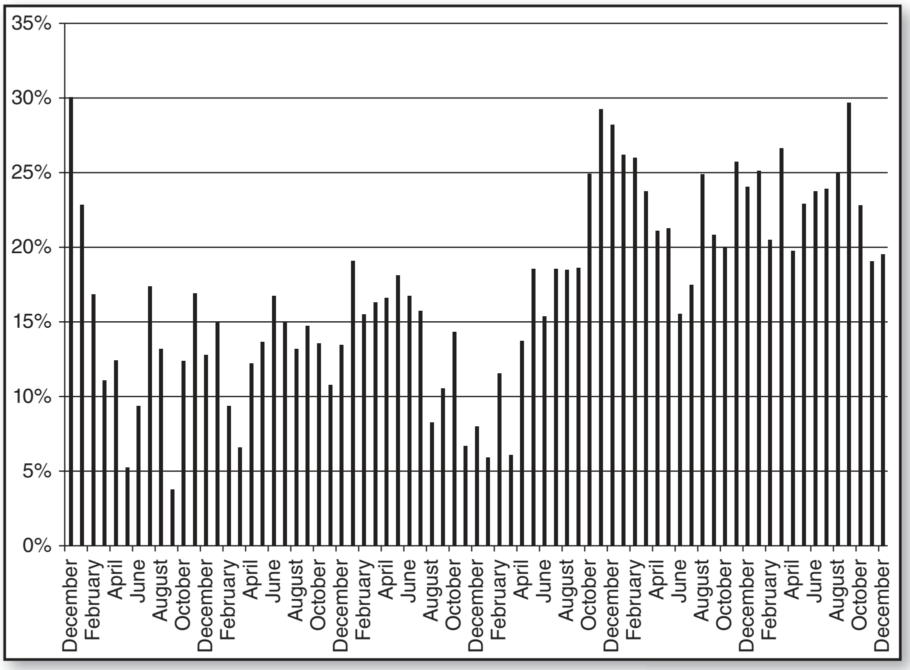

FIGURE 20.10 24-Month Rolling Return: Manager F

Underwater Curve and 2DUC Charts

The underwater chart shows the worst possible cumulative percentage loss any investor could have experienced as of the end of each month—an assumption that implies an investment started at the prior NAV peak. The low point in the NAV chart is the maximum retracement (the risk measure used in the MAR and Calmar ratios). The underwater chart, however, provides far more information because it shows not only the worst possible loss for the entire track record (the maximum retracement), but the worst possible loss as of the end of every other month in the track record as well. Figure 20.11 illustrates the underwater chart for the same two managers with widely disparate stability of returns depicted in Figure 20.6. The difference between the two could hardly be starker. Manager F's retracements are very shallow and relatively short-lived (a rise to the 0 percent level indicates a new NAV high); Manager E's retracements are both deep and protracted. The underwater chart provides an excellent visual representation of an investment's relative risk in a way that is very consistent with the way most investors perceive risk.

One shortcoming of the underwater curve is that it will understate risk for months in the early portion of the track record because there is an insufficient look-back period for a prior NAV peak. For these earlier months, there is no way of assessing a true worst-case loss representation, because a prior track record of sufficient length simply does not exist. Also, the underwater curve is constructed from the perspective of the worst cumulative loss that could have been

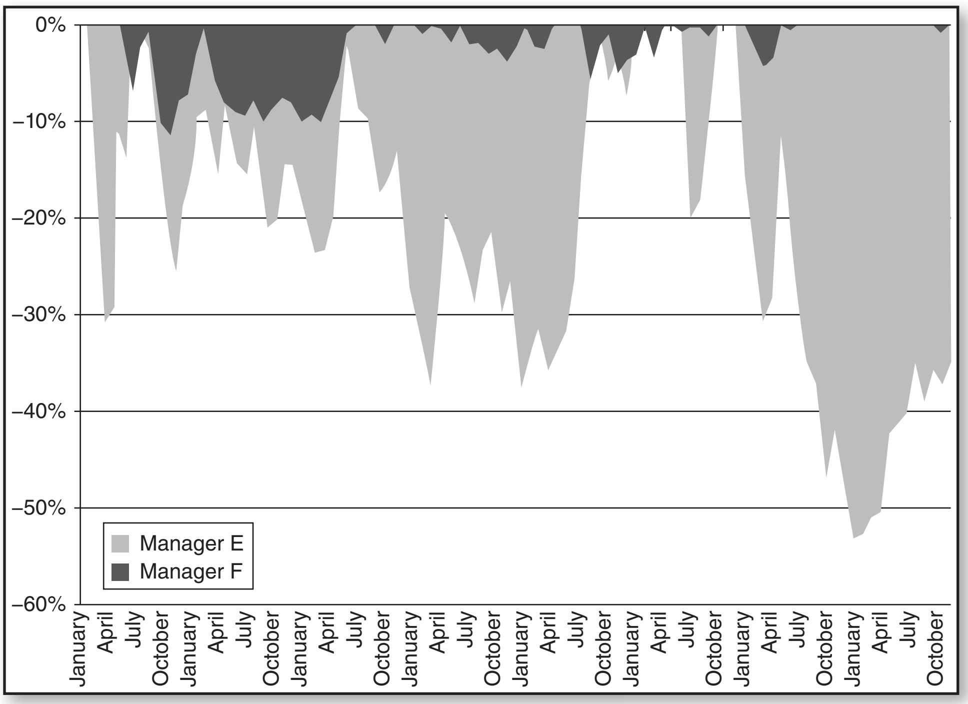

FIGURE 20.11 Underwater Curve: Manager E vs. Manager F

experienced by an existing investor. Arguably, the worst loss suffered by new investors may be an even more relevant measure. One solution to these inadequacies in the underwater curve calculation is to also consider the worst loss that could have been experienced by any investor starting in each month, assuming the investment was exited at the subsequent lowest NAV point. We can then create a two-direction underwater curve (2DUC) that for each month would show the maximum of the following two losses:

1. The cumulative loss of an existing investor starting at the prior NAV peak.
2. The cumulative loss of an investor starting that month-end and liquidating at the subsequent NAV low.

The average of all the points in the 2DUC chart would, in fact, be the risk measure used in the return retracement ratio (the average maximum retracement). The underwater excursions for Manager E become significantly more extreme in the 2DUC chart (Figure 20.12), widening from an average monthly value of 21 percent to 30 percent (the AMR). The underwater curve for Manager F remains subdued in the 2DUC chart with a still very low average value of 3 percent. The 2DUC chart implies that the average worst-case scenario for investors with Manager E is 10 times worse than with Manager F; that is a lot of extra risk for a 1.3 percent difference in the average annual compounded return. Based on performance, it would be difficult to justify choosing Manager E over Manager F, even for the most risk-tolerant investor.

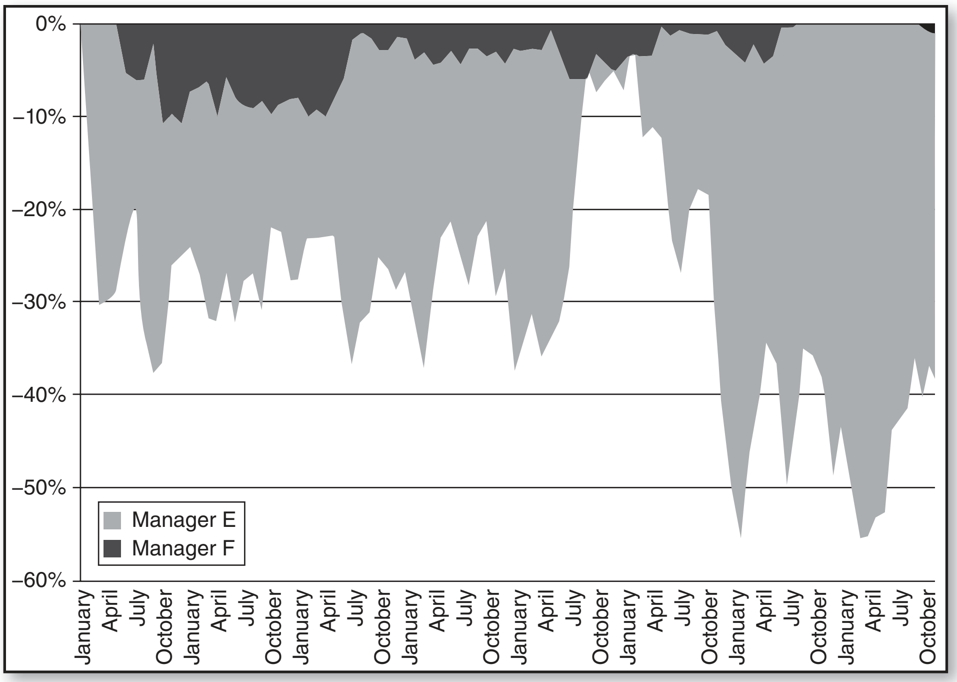

FIGURE 20.12 2DUC: Manager E vs. Manager F

Investment Insights

Many investors place too much emphasis on return. Since return can always be improved by increasing exposure (i.e., taking on greater risk), the return/risk ratio is a far more meaningful performance measure. An investment with higher return/risk and lower return than an alternative investment with the reverse characteristics can be brought up to the same higher return level with lower risk by using leverage.

The Sharpe ratio is by far the most widely used return/risk metric. The Sharpe ratio, however, penalizes upside volatility the same as downside volatility, which is not consistent with the way most investors view risk. Other return/risk measures detailed in this chapter, which focus on losses as the proxy for risk, more closely reflect the way most investors perceive risk. Investors can use Table 20.5, which summarizes the properties of different return/risk measures, to select the performance measures that best fit their criteria.

Return/risk statistics can be supplemented with the performance charts detailed in this chapter, which provide a tremendous amount of information in an intuitive and accessible format and should be at the core of any performance analysis. I recommend using the following performance charts in any manager or fund evaluation:

- An NAV chart
- Both 12-month and 24-month rolling window return charts
- A 2DUC chart

■ ■ ■

Note: Some of the statistics and chart analytics described in this chapter are my own invention and hence not yet available on any existing software. Many of these statistics and analytical charts can be accessed for free on FundSeeder.com.

PART V

FUNDAMENTAL ANALYSIS

CHAPTER 21

Fourteen Popular Fallacies, or What Not to Do Wrong

The fault, dear trader, is not in the fundamentals, but in ourselves.

(With apologies to Shakespeare)

Five Short Scenes

Scene 1

The U.S. Treasury announces a new plan to sell stockpiled gold. Not surprisingly, the market opens with near-limit losses the following day. You reason that the new gold sales will sharply increase supply and that, therefore, the market still offers a good selling opportunity, even given the decline. You are somewhat concerned about expectations for continued increasing inflation and dollar weakness but decide the gold sale will dominate market action over the near term.

After going short, the market hovers for two days and then, as you expected, breaks sharply. One week later, your trade is substantially in the plus column and convinced that you have caught a new bear market in its infancy, you resolve to hold the position as a long-term trade. The next week, however, the market begins to rally inexplicably, and your profits evaporate. Paradoxically, despite an absence of any meaningful bullish news, the rally continues and prices even surpass the levels they were at before the U.S. Treasury announcement. Your losses continue to grow, and finally you bail out, promising yourself, "That's the last time I trade on fundamentals."

Scene 2

You've done your homework and feel confident the U.S. Department of Agriculture's 50-state Hogs and Pigs report, which will be released in the afternoon, will reflect a large expansion in hog production. You anticipate that hog numbers will be up at least 7 percent over the year-ago level. Hog prices have already sold off sharply in recent weeks, but you reason that a report in line with your expectations will push prices still lower.

Although you are quite familiar with the dangers of riding a position into a major report, this is one time you cannot resist. At the report release time, eyes glued to your computer screen and your heart pounding, you read the critical figure. A smile crosses your face as you see the number. "I knew it!" you shout triumphantly. The report shows hog numbers up 8 percent.

The next day the market opens limit down, and you begin to calculate what your profits will be after three limit-down days—a conservative assumption. But before you can even finish calculating your profits, a strange thing happens: The market begins to rally. By the end of the session, hog prices are actually 100 points higher! The uptrend continues in subsequent trading sessions, and one week later you liquidate your position with a sizable loss. You feel cheated. You were right in your expectations: The report *was* bearish, wasn't it?

Scene 3

You've been long corn for three weeks and it's been one of your best trades ever. The market has moved steadily and sharply higher with export rumors flying in all directions. That evening on the news, the lead story is the official announcement of an additional large grain sale to Japan. Daydreaming, you wonder if this is the trade you will retire on.

Next morning you call your broker. "Corn is due 8 to 10 cents a bushel higher," he says. Not as good as you thought, but it will do. However, by the time corn is ready to open, the call has dropped to unchanged, and the market actually opens 2 cents lower. Several days later, corn has fallen more than 40 cents, and your profits have virtually evaporated.

Scene 4

Cattle futures have rallied to near all-time record highs. You are well aware that cattle supplies are down and expected to remain low, but upon closer examination, you have discovered that supplies were lower in a number of other past situations when prices were lower. You reason the current rally is overdone and go short.

When cattle rallies an additional 10¢/lb, you figure the market is an even better short, and add to your position. Prices are still moving higher when you finally throw in the towel on the trade.

Scene 5

You have read that sugar prices are below costs of production, a factor that seems to suggest that prices have overdone the down side. You go long. Not only does the price fail to rise, but it actually continues to slide steadily lower. You can't understand why producers continue to sell sugar at a loss. You are both confused and frustrated at the seemingly logic-defying market price action as your losses continue to mount.

■ ■ ■

These five scenes appear to provide proof that fundamental analysis just does not work. At least, that is the conclusion a great many futures traders have drawn from such experiences.

The simple truth, however, is that much that passes for fundamental analysis is either incomplete or incorrect—and frequently both. The trader who ignores fundamentals completely is almost certainly better off than the trader who uses fundamentals incorrectly. However, this in no way alters the fact that good fundamental analysis is a useful, and even powerful, tool.

Before turning to how to do things right, it is essential to first cover what not to do wrong. We begin by exploring 14 common fallacies in fundamental analysis. Incidentally, these fallacies do not represent mistakes made solely by the novice trader. In fact, virtually all these errors have been repeated numerous times in the most respected financial news outlets and in myriad commodity research publications. There is no significance to the order of the list.

The Fourteen Fallacies

1. Viewing Fundamentals in a Vacuum

"The fundamentals are bearish" is often thought to be synonymous with an abundant supply situation. Such an interpretation might seem plausible, but it can lead to inaccurate conclusions.

For example, assume the sugar market is trading at 30¢/lb. and in transition from tightness to surplus. Given this scenario, the fundamentals can indeed be termed "bearish," and an expectation of lower prices would be reasonable. Assume that prices begin to move lower. Are the fundamentals still bearish at 25¢? Very likely. At 20¢? Maybe. At 15¢? At 10¢? At 5¢? The point is that at some price level, the fundamentals are no longer bearish, no matter how large the projected supply.

In fact, it is entirely possible that the fundamentals could be bullish in a surplus situation if prices have overdone the downside—a situation that is far from infrequent. Thus, fundamentals are not bullish or bearish in themselves; they are only bullish or bearish relative to price. The failure of many analysts to realize or acknowledge this fact is the reason why the fundamentals are so often termed *bullish* at market tops and *bearish* at market bottoms.

2. Viewing Old Information as New

Financial news outlets frequently report old information and new information in much the same manner. For example, a story with the headline "World Cotton Production Projected to Rise 10 Percent" may sound very bearish. However, what the story is not likely to indicate is that this may be the fourth or fifth such estimate released. Very likely, the previous month's estimate also projected an approximate 10 percent increase in world production. For that matter, the previous month's estimate might have forecast a 12 percent increase, and the current estimate actually represents a price-constructive development. The main point to keep in mind is that much information that sounds new is actually old news, long discounted by the market.

3. One-Year Comparisons

The use of one-year comparisons is fairly widespread, probably because it offers a simple means of instant analysis. This approach is overly simplistic, however, and should be avoided. For example, consider the following market commentary: "The December Hogs & Pigs Report indicates that large pork supplies are around the corner. Market hogs on all farms are up 10 percent. The projected 10 percent increase in hog slaughter should push prices lower. . . ." Although this type of analysis could be right on target in some situations, it will be susceptible to error if used consistently.

Sharp-eyed readers may already be citing fallacy number 1—that is, large supplies do not necessarily imply lower prices, since the market may already be discounting such a development. However, some additional potential errors pertain specifically to the one-year comparison. First, just because the December report indicated a 10 percent increase in hog numbers does not mean it implies large supplies. Perhaps hog numbers were extremely low the previous year. Second, the relationship between hog slaughter and market hogs can vary significantly. It is possible that the preceding year the ratio of slaughter to market hogs was abnormally high. In this case, a 10 percent increase in market hogs would imply a smaller increase in slaughter. Although one-year comparisons can be used sparingly for illustrative purposes, they should never represent the sole basis of fundamental analysis.

4. Using Fundamentals for Timing

If this list of fallacies were ordered on the basis of frequency of occurrence, this item would be a strong contender for the number 1 spot. Fundamental analysis is a method for gauging what price is right under given statistical conditions and can be used in constructing annual, quarterly, and in some instances monthly price projections. However, it is ludicrous to attempt to boil supply-demand statistics down to the point at which they provide an instantaneous price signal, which is exactly what some traders do when they rely on fundamentals for timing.

Trading on the basis of market websites, newspaper articles, and newswire stories, falls into this category. It is no surprise that speculators who base their trades on such items are usually spectacularly unsuccessful. The only major exception are those traders who use this type of information in a contrary way, such as viewing the failure of the market to rally after the release of a bullish newswire story as a signal to go short.

The fundamental researcher must also guard against the natural instinct of wanting to take a market position right after completing an analysis that indicates either an underpriced or overpriced situation. The market is not aware of the timing of a researcher's personal price discovery. Even if the analysis is correct, the right time may be three weeks or even three months off. In short, for purposes of timing, even the fundamental analyst should use some form of technical input.

5. Lack of Perspective

Assume the following scenario: Scanning the financial pages one day, you notice the following headline: "Government Officials Estimate 10,000 Head of Cattle Killed in Recent Midwest Winter Storm." Does such a large production loss suggest a major buying opportunity? Wait a minute. What large production loss? Ten thousand cattle might sound like a very big number if you were to picture them on your front lawn, but viewed in terms of a total U.S. cattle population of about 90 million head (and many times that globally), the loss does not even equal the proverbial drop in the bucket.

This example is based on supply, but cases involving domestic consumption or exports could be illustrated just as easily. In each instance, the same question should be asked: How important is the event (e.g., production loss, new export sales) in terms of the total picture?

6. Ignoring Relevant Time Considerations

True or false: Higher grain prices imply higher meat prices. No cheating—think before reading on.

Actually, this is not a fair question, because the answer depends on the time frame. Most people would probably answer true, since rising grain prices do suggest increased costs of production for feedlot operators, a development that would lead to reduced meat production and higher meat prices. (Cost of production is itself a primary source of misconception and is discussed separately.) However, this reasoning is true only for the very long run (2½ years plus).

Over the short to intermediate term—the time frame that is really of primary concern to futures traders—the effect might be exactly the opposite. If high grain prices are effective in influencing cattle feeders to reduce production, the preliminary impact will be increased marketings and lower prices as a result of breeding herd liquidation. Higher grain prices might reduce the weight to which cattle are fed, but this effect is relatively minor. Increased feeding costs would only imply a shift in the flow of supply (since cattle gain weight more slowly on grass) rather than a change in total actual supplies over the longer run.

In the world of economics, the cause-and-effect relationship is not necessarily instantaneous. In some cases, an event will trigger a very quick price response; in other instances, such as the cattle situation, the effect will not occur for many years.

7. Assuming That Prices Cannot Decline Significantly Below the Cost of Production

No matter how many times this old saw is disproved by actual events, it never seems to be laid to rest. The cost of production is not—repeat, not—a price-supporting factor, especially for nonstorable commodities.

Once a commodity is produced, the market does not care about the cost of production. Prices will be determined on the basis of existing supply and demand. If prices fall to the cost of production and there is still a surplus, prices will continue to decline until an equilibrium price level is reached.

Why should producers sell a commodity below the cost of production? The fact is they don't have much choice. Agricultural markets are highly competitive, with literally thousands of sellers. Consequently, any individual is powerless to pass on production costs to the marketplace. Instead, individuals must accept the price that the market will bear. After all, a low price is better than no price.

Of course, an unprofitable situation will lead to production cutbacks, but this will not happen overnight. The minimum time lag might be one year, but in many instances, it will take several years before prices below the cost of production actually result in reduced output. In this sense, fallacy number 7 is a corollary of fallacy number 6—ignoring relevant time considerations. Many commodity markets have witnessed periods in which prices have fallen and stayed below cost of production for years at a time. Keep this empirical reality in mind the next time you read a recommendation to purchase a commodity because it is at or below the cost of production.

8. Improper Inferences

Fallacy number 8 might be best explained by citing some examples. First, cattle-on-feed numbers do not necessarily provide an indication of potential future slaughter. Reason: cattle on feed do not include grass-fed cattle. As long as grass-fed cattle account for a stable percentage of total slaughter, there is no problem. But if the percentage varies widely over time (as has tended to be the case), the straightforward use of cattle-on-feed numbers to predict slaughter can lead to a totally erroneous conclusion. If, for instance, high feed prices influence a shift toward increased grass feeding of cattle, the total number of cattle could be higher, even if the cattle-on-feed figure shows a significant reduction.

Much market analysis and commentary naively ignores the preceding complication in projecting cattle slaughter. How bad is this error? Table 21.1 shows the relationship between percentage changes in cattle on feed and total slaughter. There is a great deal of variability between the two sets of figures. In fact, from 1995 through 2014 there were 34 quarters when the deviation in percentage changes between cattle on feed and total slaughter exceeded 5 percent and seven quarters with deviations greater than 10 percent! It is not an overstatement to say that one can achieve *far* more accurate slaughter projections using the naive assumption that slaughter in any given quarter will equal the corresponding previous year's level. This is a clear example of no information being far preferable to incorrectly used information.

TABLE 21.1 Percentage Changes in Cattle on Feed Numbers Versus Percentage Changes in Slaughter

Quarter	Cattle on Feed as Percentage of Previous Year	Cattle Slaughter as Percentage of Previous Year	Discrepancy between Two Percentages[a]
Jan-2015	94.48%	100.98%	6.50%
Oct-2014	91.17%	99.49%	8.31%
Jul-2014	91.72%	97.61%	5.89%
Apr-2014	94.14%	99.53%	5.39%
Jan-2014	94.77%	94.79%	0.02%
Oct-2013	97.01%	92.31%	–4.70%
Jul-2013	99.85%	96.81%	–3.05%
Apr-2013	100.19%	95.01%	–5.18%
Jan-2013	96.95%	94.37%	–2.58%
Oct-2012	98.66%	97.40%	–1.26%
Jul-2012	95.37%	102.66%	7.28%
Apr-2012	96.18%	102.00%	5.82%

TABLE 21.1 (*Continued*)

Quarter	Cattle on Feed as Percentage of Previous Year	Cattle Slaughter as Percentage of Previous Year	Discrepancy between Two Percentages[a]
Jan-2012	96.53%	103.02%	6.49%
Oct-2011	97.00%	104.86%	7.85%
Jul-2011	99.84%	103.74%	3.90%
Apr-2011	99.54%	104.92%	5.39%
Jan-2011	101.85%	104.83%	2.99%
Oct-2010	105.02%	102.91%	–2.10%
Jul-2010	102.74%	103.26%	0.52%
Apr-2010	100.89%	96.46%	–4.42%
Jan-2010	102.36%	97.99%	–4.37%
Oct-2009	100.78%	100.57%	–0.22%
Jul-2009	96.14%	94.73%	–1.41%
Apr-2009	94.99%	95.53%	0.54%
Jan-2009	96.44%	92.90%	–3.53%
Oct-2008	95.30%	94.97%	–0.33%
Jul-2008	101.84%	95.88%	–5.96%
Apr-2008	102.58%	100.34%	–2.24%
Jan-2008	101.42%	101.03%	–0.39%
Oct-2007	103.22%	96.33%	–6.89%
Jul-2007	99.60%	98.76%	–0.84%
Apr-2007	100.25%	98.58%	–1.67%
Jan-2007	103.97%	101.44%	–2.53%
Oct-2006	103.73%	108.61%	4.89%
Jul-2006	102.94%	104.60%	1.66%
Apr-2006	106.22%	108.64%	2.41%
Jan-2006	103.25%	104.47%	1.22%
Oct-2005	100.45%	99.81%	–0.64%
Jul-2005	101.69%	102.57%	0.88%
Apr-2005	97.22%	100.99%	3.77%
Jan-2005	96.43%	100.41%	3.98%
Oct-2004	98.28%	102.73%	4.45%
Jul-2004	87.31%	101.96%	14.65%
Apr-2004	90.09%	100.33%	10.24%
Jan-2004	94.33%	105.58%	11.26%
Oct-2003	91.22%	98.05%	6.83%
Jul-2003	103.14%	94.62%	–8.51%
Apr-2003	103.37%	92.45%	–10.92%
Jan-2003	99.29%	91.60%	–7.70%
Oct-2002	100.62%	93.63%	–6.99%

(*Continued*)

TABLE 21.1 (*Continued*)

Quarter	Cattle on Feed as Percentage of Previous Year	Cattle Slaughter as Percentage of Previous Year	Discrepancy between Two Percentages[a]
Jul-2002	103.08%	95.24%	–7.84%
Apr-2002	101.37%	100.47%	–0.90%
Jan-2002	98.93%	98.03%	–0.91%
Oct-2001	100.65%	100.99%	0.34%
Jul-2001	97.13%	105.89%	8.76%
Apr-2001	98.22%	102.87%	4.65%
Jan-2001	94.41%	102.81%	8.40%
Oct-2000	98.64%	107.20%	8.56%
Jul-2000	99.18%	108.61%	9.43%
Apr-2000	100.26%	107.58%	7.33%
Jan-2000	103.09%	107.57%	4.48%
Oct-1999	102.15%	104.96%	2.81%
Jul-1999	102.89%	104.30%	1.41%
Apr-1999	102.03%	102.63%	0.60%
Jan-1999	100.62%	95.63%	–4.99%
Oct-1998	98.42%	97.83%	–0.59%
Jul-1998	97.99%	102.27%	4.28%
Apr-1998	96.69%	97.27%	0.58%
Jan-1998	97.54%	105.65%	8.11%
Oct-1997	99.57%	112.69%	13.12%
Jul-1997	101.46%	114.26%	12.80%
Apr-1997	97.00%	105.90%	8.90%
Jan-1997	99.19%	102.05%	2.86%
Oct-1996	100.12%	96.94%	–3.17%
Jul-1996	98.32%	85.05%	–13.27%
Apr-1996	105.91%	99.50%	–6.42%
Jan-1996	106.58%	107.92%	1.34%
Oct-1995	103.04%	101.63%	–1.41%
Jul-1995	105.14%	106.00%	0.85%
Apr-1995	105.48%	100.17%	–5.31%
Jan-1995	103.14%	94.61%	–8.53%
		Avg. (abs)	4.73%
		Med. (abs)	4.40%
		Max. (abs)	14.65%
		Min. (abs)	0.02%
Qtrs. w/ abs discrepancy ≥ 5%			34
Qtrs. w/ abs discrepancy ≥ 10%			7

[a]Column 2 minus column 3 percentages.

Another example of an improper inference is provided by the projection of production from acreage figures. A given percentage change in acreage does not necessarily imply a similar change in production (even assuming equivalent yields). For most crops, the distribution of production is a critically important variable. For example, average cotton yields in some states, such as California, are approximately three times as high as average yields in other states, such as Texas. Although considerably more time consuming, production projections should be based on a breakdown of acreage by area (region or state) rather than on a total acreage figure.

9. Comparing Nominal Price Levels

It is inaccurate to compare current prices with the actual recorded prices of previous years. In drawing comparisons with past seasons, it is necessary to adjust historical prices for inflation. Even though U.S. inflation has been subdued since the mid-1980s, over broad periods of time, even low inflation can have a significant cumulative effect. Moreover, in the future, higher inflation levels could recur, making this factor a critical consideration.

As an example, assume that an exhaustive survey of the statistical data for commodity *x* in past years indicates that 1997 and 2003 were very similar to the current season in terms of overall fundamentals. Does this observation imply that current-season prices will be about in line with the price levels of 1997 and 2003? Of course not. In real dollar terms, the prices may be roughly equivalent, but because of the impact of inflation, current nominal prices are likely to be higher.

Inflation cannot be considered in a vacuum, however. For example, a protracted downshift in demand for most physical commodities (resulting from reduced inventory requirements) beginning around 1980 provided a counterbalancing force to inflation. Because demand is very difficult to quantify—as will be discussed in detail in Chapter 22—the net effect is that inflation-adjusted forecasts can be biased to the high side. In other words, ironically, in some cases it is possible that a naive analyst who ignores both demand shifts and inflation adjustments may derive a more accurate forecast than the analyst who adjusts for inflation. Such accidental accuracy is likely to be a temporary phenomenon. The correct procedure would be to incorporate inflation adjustments in the model and then infer and include demand shifts in the model as well.

10. Ignoring Expectations

Markets often place greater emphasis on expectations for the following year (or season) than on prevailing fundamentals. This pattern is especially true in transitional periods when the supply situation is moving from surplus to tightness, or vice versa.

The 1990 wheat market provided an excellent example. In the 1989–1990 season, the winter wheat crop proved very disappointing because of below-average yields. As a result, carryover stocks (measured as a percent of utilization) fell to their lowest level in 15 years. Moreover, winter wheat seedings for the 1990 crop increased only slightly, thereby seeming to suggest the extension of a tight supply situation into the new season.

Despite the apparent bullish scenario, wheat prices moved steadily and sharply lower from the very beginning of 1990. This price slide cannot be explained in terms of prevailing fundamentals, but only in terms of expectations. As the year progressed, it became increasingly evident that the

1990–1991 hard red winter wheat crop would result in extremely good yields. As it turned out, the yield of the 1990–1991 winter wheat crop increased by an imposing 16 percent over the previous season's level, and the percentage of planted acreage harvested rose from 75 percent to 88 percent. As a result of excellent yields and sharply lower acreage abandonment, 1990–1991 winter wheat production increased by a huge 39 percent, despite the fact that planted acreage was only marginally higher, and carryover stocks returned to comfortable levels.

Although the fundamental transition just described was reflected by data available after mid-spring 1990, during early 1990 such changes would have fallen into the category of expectations. Thus, price action in the wheat market during the first half of 1990 provided a classic example of expectations dominating prevailing fundamentals.

11. Ignoring Seasonal Considerations

Almost every commodity exhibits one or more seasonal patterns. Ignoring seasonal factors can easily lead to the misinterpretation of fundamental data. For example, a 5 percent increase in hog slaughter during the fourth quarter relative to the third-quarter level would actually be indicative of a trend toward reduced production—not expanded production. The explanation behind this apparent paradoxical statement lies in the fact that hog production is highly seasonal. Producers breed hogs so that the largest pig crop is born during the spring and the smallest in winter. Because it takes approximately six months for hogs to reach market weight, slaughter tends to be heaviest during the fall and lightest during the summer. Thus, it is essential to adjust for the seasonal production pattern in drawing slaughter comparisons between the current period and the preceding month or quarter.

Comparisons of production and consumption figures with the corresponding figures of previous years obviously do not require any consideration of seasonal factors. However, if comparisons of fundamental data involve different time periods during the year, it is essential to examine historical data carefully for possible seasonal behavior and to make any necessary adjustments.

12. Expecting Prices to Conform to Target Levels in World Trade Agreements

The history of commodities is replete with examples of world trade agreements that totally failed to achieve their stated goals. Trade agreements typically attempt to support prices through export controls and stockpiling plans. Although these provisions provide some underlying support to the market and occasionally even spark temporary rallies, they are usually not sufficiently restrictive to maintain prices significantly above equilibrium levels for any extended period of time. The International Sugar Agreement and the International Cocoa Agreement are two examples of world trade agreements that ultimately failed to support prices above the lower end of their respective stated target ranges (in the years when these agreements attempted to support prices; they no longer even attempt to do so). Perhaps the most effective price-supporting organization has been the Organization of the Petroleum Exporting Countries (OPEC), but even the oil cartel has frequently seen prices fall below their target level—often by a wide margin.

It should be noted that world trade agreements are even more impotent in terms of restraining a price advance. In the case in which prices approach the upper end of a target range, the most powerful action that any agreement could take would be the elimination of all restrictions—in other words, a return to a free market.

13. Drawing Conclusions on the Basis of Insufficient Data

Sometimes it is virtually impossible to construct a fundamental forecasting model for a market because of a lack of sufficient comparative historical data. A perfect case in point was provided in the August 1972 issue of *Commodities* magazine (now called *Modern Trader*), which ran a detailed study of fundamentals in the cotton market. The article ultimately came to the valid conclusion that only two seasons since 1953 could truly be termed *free markets.* As the article explained, during the 1950s and 1960s, government programs had maintained cotton prices above the levels that would have been realized had prices been determined by the interaction of supply and demand. So far, so good.

The proper and very worthwhile conclusion would have been that existing data were insufficient to permit the use of fundamentals in forecasting prices. After all, how can you interpret the price implications of a projected statistical balance if there are only two previous years to use as a comparison?

Unfortunately, the author went on to sketch an entire set of price forecasting conclusions on the basis of admittedly very limited relevant information. Quoting the first item, "Final stock levels under 3½ million bales imply a very tight supply situation and suggest a likelihood of a price rise well above 30¢ in such seasons."

Although this statement certainly proved true, by implication it severely understated the upside potential in the cotton market. Only a little more than one year after the article was published, cotton prices reached an all-time peak of 99¢/lb. Incidentally, I was the author of that article.

14. Confusing the Concepts of Demand and Consumption

Demand is probably one of the two most misused words in futures literature and analysis (parameter being the other; see Chapter 19). The confusion between demand and consumption is not a matter of semantics; the two terms represent very different concepts, and their frequent interchangeable use leads to many major analytical errors. An adequate explanation of this statement requires a diversion into a short review of basic supply-demand theory, which is the subject of the next chapter.

At this point, it might be instructive to return to the five scenes depicted at the beginning of this chapter to try to determine which of the 14 fallacies were responsible for the incorrect trading conclusions. Note that each scene reflects two or more fallacies. The answers can be found in Table 21.2.

TABLE 21.2 Fallacies Committed in the Five Scenes

Scene	Fallacies[a]
1	4, 5, 10
2	1, 3, 4
3	2, 4
4	9, 10
5	7

[a]The inclusion of additional items is not necessarily incorrect. Other fallacies might also be applicable (e.g., fallacy number 1 in any of the scenes), but are not listed because the text provides insufficient information to make such a determination.

CHAPTER 22

Supply-Demand Analysis: Basic Economic Theory

There are in the fields of economics no consistent relations, and consequently, no measurement is possible.

—Ludwig Edler von Mises

Supply and Demand Defined

Supply curves slope upward, meaning more is offered to the market at higher prices (Figure 22.1).[1] Assuming that the time unit shown on the horizontal axis in Figure 22.1 equals one season, the supply that can be offered to the market will be limited to total production plus stocks, regardless of the price. At high prices, however, producers will be willing to hold smaller inventories and therefore offer greater quantities to the market. Conversely, at lower prices, producers will prefer to store

[1] The supply and demand curves in this section are drawn as straight lines for simplicity of exposition. It also seemed desirable to avoid the unnecessary digression of discussing the factors that determine the precise shapes of these curves. Although the straight-line assumption may often be adequate within normal boundaries, supply and demand curves will not be linear over the entire price range. For example, as prices rise and the quantity consumed declines, it will usually take greater and greater increases in price to induce a given further reduction in the amount consumed. As another example, over the short run, at some point the supply curve must begin to rise asymptotically, since the supply offered to the market cannot exceed the existing total supply (i.e., stocks plus current production).

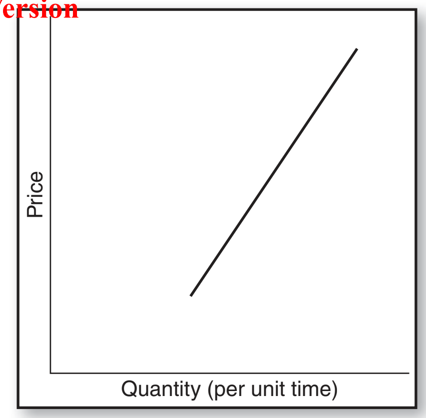

FIGURE 22.1 Supply Curve

larger quantities rather than marketing their goods at prevailing depressed levels. The slope of the supply curve will reflect this tradeoff between the options of sale and storage.[2]

For perishable (e.g., eggs, potatoes[3]) or nonstorable (e.g., cattle, hogs) commodities, supply is approximately fixed and can be represented by a vertical line (Figure 22.2). For example, if a supply curve is drawn for the hog market for a time unit of one-half year, the amount offered to the market during that period will be relatively independent of market prices. Low prices will not reduce the quantity supplied, because once hogs reach market weight, with the exception of temporary delays, producers have little choice but to bring those hogs to market, regardless of the price. However, since there is a lag of nearly one year between producers' breeding decisions and the time that resulting

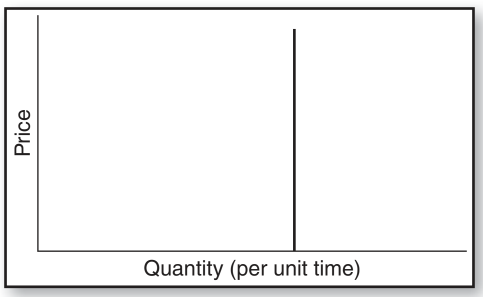

FIGURE 22.2 Fixed Supply

[2] For longer time units (e.g., 10 years), the supply curve will also reflect the potential for an expansion in production beyond current levels. For example, high prices may encourage shifts in acreage to the high-priced commodity and increased usage of fertilizer in new crops. From the vantage point of futures trading, however, it is most useful to limit the discussion of supply and demand to short time units (i.e., season or fraction of a season).

[3] These commodities are no longer traded as futures markets, but provide perfect illustrations of perishable goods.

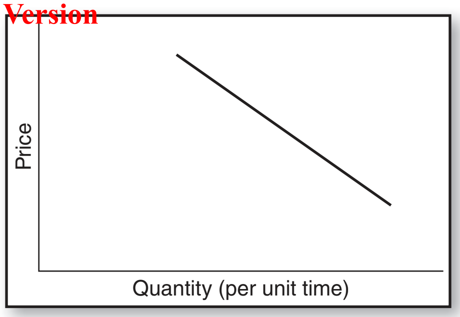

FIGURE 22.3 Demand Curve

offspring reach market weight, high prices cannot induce an increase in the quantity supplied. In fact, if anything, the supply curve in such a market exhibits a perverse behavior; that is, high prices will reduce the quantity supplied. The reason is that high prices will influence producers to withhold hogs from the market for breeding, thereby reducing current supplies. However, for simplicity's sake, we will assume a vertical supply curve in the case of perishable or nonstorable commodities.

Demand can be defined as a schedule of the various quantities of a commodity that will be consumed at each price level. In a sense, demand is a barometer of consumer buying pressure. Demand curves slope downward, meaning more will be demanded at lower prices (Figure 22.3).

Elasticity of demand can be defined as the percentage increase in the amount demanded divided by the percentage decrease in price. If the demand for a commodity is inelastic, it means that a relatively large percentage change in price will only induce a small percentage change in the amount demanded. Figure 22.4 presents illustrations of elastic and inelastic demand curves.[4]

The elasticity of demand is primarily determined by two basic factors:

1. **Availability of substitutes.** The elasticity of demand will vary directly with the availability of substitutes. For example, the demand for salt is highly inelastic, but the demand for a given brand of salt is very elastic.
2. **Percentage of total income spent on the good.** The elasticity of demand will vary directly with the percentage of expenditures allocated to a good. For example, the demand for automobiles is far more elastic than the demand for salt, even though there are no close substitutes for either item.

Generally speaking, the demand curves for most commodities tend to be inelastic; that is, a given percentage change in price will induce a smaller opposite percentage change in the amount demanded. This is a significant consideration, since prices of goods with inelastic demand curves are more subject to wide price swings in times of shortage.

[4] Elasticity is not constant along each demand curve. Elasticity is a concept that relates to a given point, not to the entire curve. As we move rightward along a line or demand curve (in both the elastic and inelastic cases), the elasticity of demand will decrease, since any given change in price will represent a larger percentage change, and will influence the same absolute, but smaller percentage, change in the quantity demanded. In other words, as can be verified in Figure 22.4, rightward movement along the demand curve will increase the denominator and decrease the numerator of the elasticity of demand.

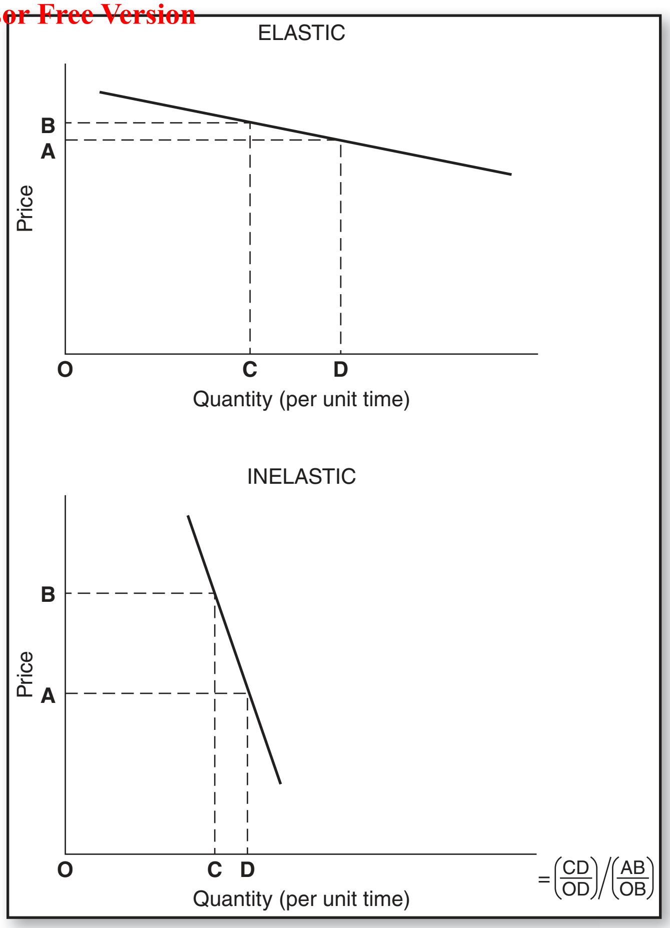

FIGURE 22.4 Elasticity of Demand

The Problem of Quantifying Demand

As all students of Economics 101 know, price is determined by the intersection of the supply and demand curves (Figure 22.5). However, there is one major problem in using supply-demand analysis to project prices: Demand is not readily quantifiable—that is, there is no way of determining how much will be consumed at any given price level. Whereas in most cases supply can be either approximately fixed as in the case of perishable and nonstorable commodities, or at least roughly estimated using production and stock statistics,[5] demand is entirely intangible. It is hardly feasible to query all

[5] The precious metal markets provide an important exception. See the section, Why Traditional Fundamental Analysis Doesn't Work in the Gold Market, at the end of this chapter.

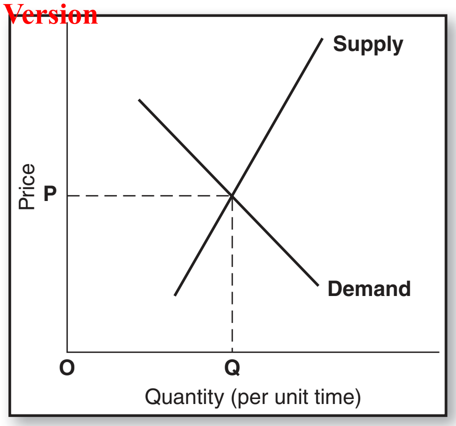

FIGURE 22.5 Equilibrium

potential consumers as to the amount of a good they would purchase at various price levels. Even if a sampling procedure were used—presumably an impractical and prohibitively expensive approach for the analyst—there is no reason to assume that consumers could even describe their demand curves.

The only theoretically acceptable means of quantifying demand is to infer demand curves through a detailed analysis of historical consumption and price data. Although this is an easy task if demand is relatively stable, unfortunately, it is either difficult or impossible if demand is subject to frequent wide shifts.

Understanding the Difference between Consumption and Demand

Perhaps the most commonly employed solution to the problem of quantifying demand is the use of consumption as a proxy for demand. This approach, however, has one major drawback: It is totally incorrect. The synonymous use of consumption and demand represents a confusion of two entirely different concepts. *Consumption* is the amount of a good used and is determined by price, which in turn is determined by supply and demand factors. *Demand* refers to the amount of a good that will be used at any given price level and, along with supply, determines price.

An increase in demand means that more will be consumed at any given price level (Figure 22.6). Factors that might affect demand include disposable income, consumer tastes, and the price of substitute goods but, by definition, not price. For most commodities, a rise in disposable income will result in an increase in demand; that is, at each given price, more will be consumed than before. A price decline will lead to increased consumption, showing movement along the same demand curve, but it does not imply anything about demand. In other words, all else being equal, the same amount will be consumed at each given price level unless there is a change in demand.

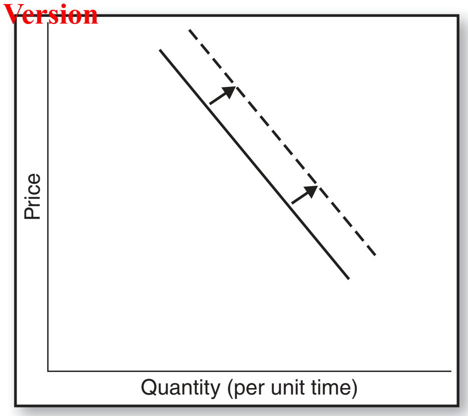

FIGURE 22.6 Increase in Demand

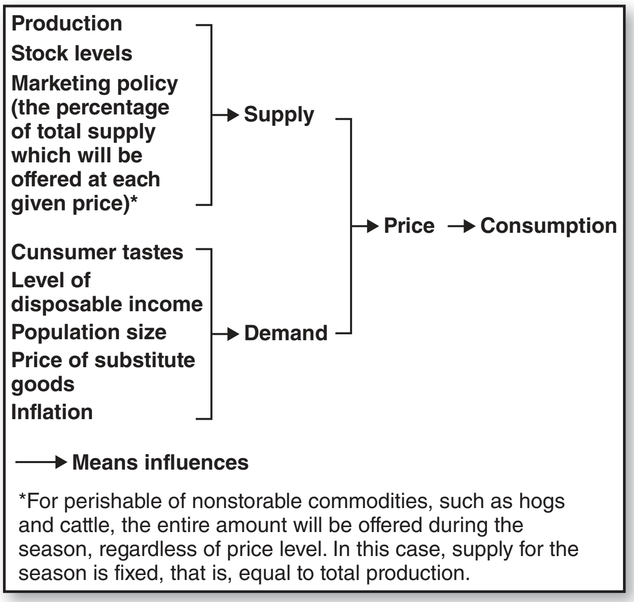

FIGURE 22.7 Supply-Demand Interaction

Figure 22.7 summarizes the relationship between demand and consumption. Consumption (i.e., the amount consumed) is directly dependent on price, where price is determined by the interaction of supply and demand. The key point to keep in mind: Consumption is a consequence of price, not a determinant of price. Thus the concept that consumption mirrors demand is totally erroneous; consumption is determined by *both* supply and demand.

In fact, for perishable and nonstorable commodities, consumption primarily reflects supply, not demand. For example, assume that pork consumption has increased sharply. Does this mean that pork demand has suddenly improved dramatically? Absolutely not. The consumption increase is merely the result of increased hog slaughter. Recalling that the supply curve for hogs (and therefore pork) can be approximated by a vertical line, Figure 22.8 demonstrates the consumption level will be determined by supply and will be the same, no matter which demand curve prevails. Thus, an increase in consumption would merely reflect an increase, or rightward shift, in supply—a bearish development—and not an increase in demand, which would be a bullish development.

It is entirely possible for a demand increase and a consumption decrease to occur simultaneously. Figure 22.9 illustrates how this is possible for both the variable supply and fixed supply cases. At the

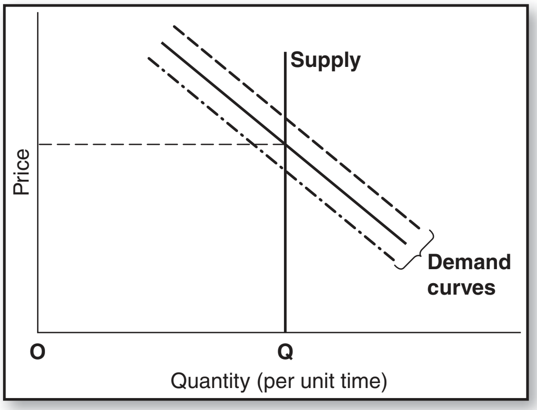

FIGURE 22.8 Consumption Reflects Supply (in Fixed Supply Case)

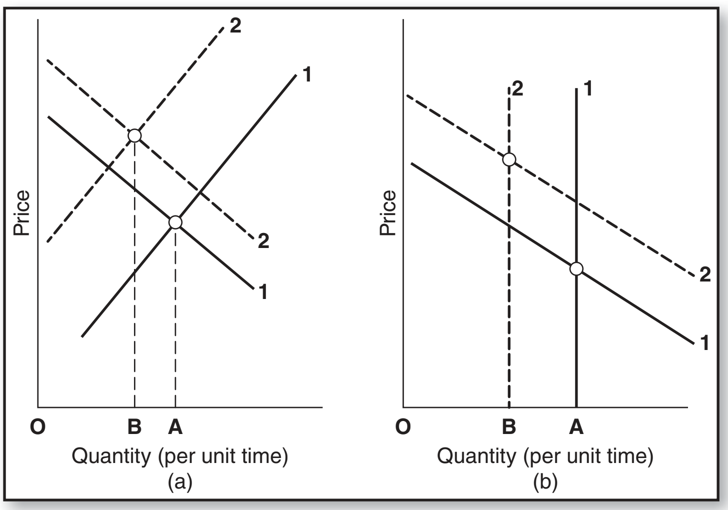

FIGURE 22.9 Higher Demand, Lower Consumption: Variable Supply (a) and Fixed Supply (b)

start, in period 1, the equilibrium consumption level is at point A. Although demand increases in period 2, the equilibrium consumption level declines to B as a result of the decline in supply.

Even the U.S. Department of Agriculture (USDA), one of the nation's leading employers of economists, has misused the term *demand*. The popularly termed *supply-demand* reports are in reality supply-disappearance reports (with *disappearance* defined as total domestic consumption plus exports).

Quite frequently, when the USDA changes its estimates for items that are sometimes discussed under the label of "demand" (domestic consumption, exports),[6] the revision reflects a change in supply, not demand. For example, if the projected carryover for a commodity is already at estimated minimum pipeline requirements, a reduced production forecast will mean that the USDA has to lower either the domestic consumption estimate, the export estimate, or both. Otherwise, the USDA might find itself in the absurd position of projecting a near-zero or even negative carryover. However, the key point is that such revisions do not imply that demand has been reduced—a bearish conclusion—but rather that high prices will ration scarce supplies, thereby resulting in reduced usage.

The Need to Incorporate Demand

Because of the difficulties involved in quantifying demand, there is often a temptation to concentrate solely on supply factors in constructing a fundamental price-forecasting model. This can be a grave error, because a demand shift can often be the dominant force in a major price move. The 1980–1982 copper market provided a classic example of the dangers of ignoring demand. To focus in on this example, we will examine a 21-year segment of the copper market (1973–1993) that contains three major bull-bear price cycles, with the bear phase of the middle cycle containing the 1980–1982 market that is the center of our attention.

Consider the following copper price-forecasting model:

$$P = f\left(\frac{S}{C}\right)$$

where P = average deflated copper price during the period
S = copper stock level (U.S. plus foreign refined copper stocks)
C = copper consumption level during the period (annualized refined copper deliveries, United States plus foreign)

$f(\ )$ is read as "is a function of," which basically means "is dependent on."

At surface glance, this model seems reasonably plausible. In essence, the model implies that prices will be low when copper stocks are large relative to the usage level and high in the reverse case. This model certainly seems logical enough. Figure 22.10, which illustrates the relationship between copper prices and the stock/consumption ratio during a 21-year segment (1973–1993) that is centered near the 1980–1982 bear market that we wish to focus on, appears to confirm this expected market behavior. The strong inverse correlation between the stock/consumption ratio and copper prices is

[6] USDA report tables, however, correctly label these items as components of "disappearance."

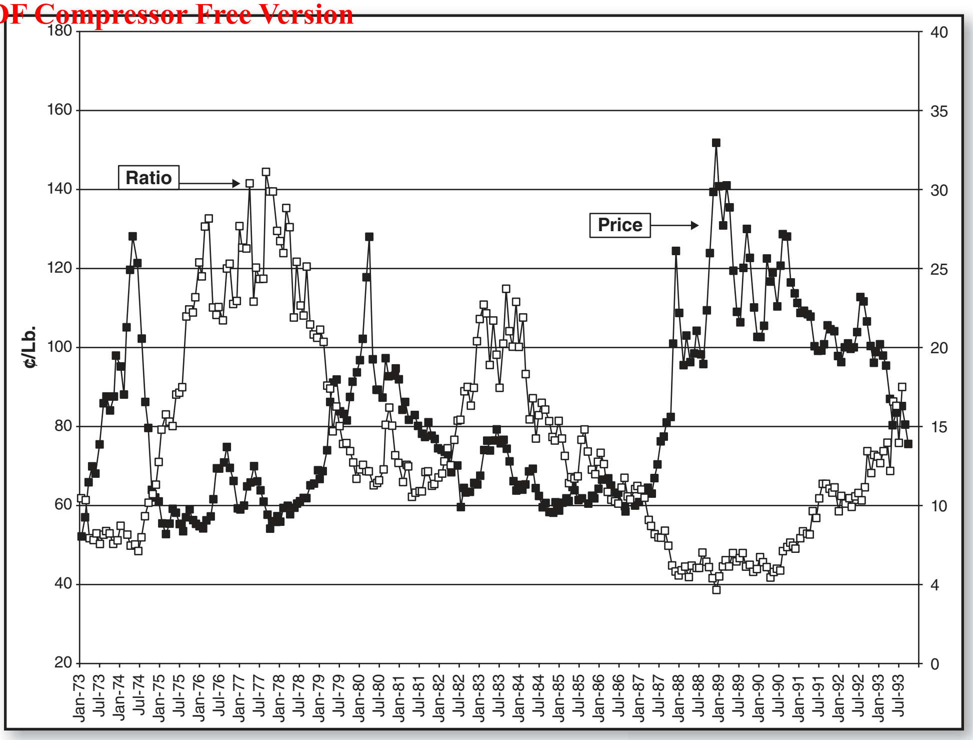

FIGURE 22.10 Average Monthly Copper Nearest Futures Price vs. Copper Stock/Consumption Ratio

broadly evident across the entire period shown. However, note the seemingly puzzling 1980 to mid-1982 price behavior. During this period, prices plunged dramatically despite a slide in the stock/consumption ratio to a major low. How can this counter-to-expected price action be explained?

There is no mystery. Although the stock/consumption ratio is an important price-influencing factor, it only reflects supply. The apparent paradoxical behavior from 1980 to mid-1982 is explained by the fact that the model does not incorporate demand. During this period, the anticipation and ultimate realization of a severe recession combined with high real interest rates (interest rate minus inflation rate) drastically reduced the inventories users wished to hold at each given price level. In other words, there was a sharp downward shift in the demand curve. This crucial fundamental development simply could not be reflected by the model just described.

The moral is that it is always necessary to take demand into account. The next section discusses several methods for incorporating demand in the price-forecasting model. But even when this type of analysis is not possible, demand must still be considered. If demand is not part of the model because of the inherent difficulties in quantifying demand, then the analysis should be divided into two steps:

1. Model projection
2. Informal evaluation of the potential impact of demand factors

Possible Methods for Incorporating Demand

How can the problem of nonquantifiable demand be circumvented? The answer depends on the market. The following types of markets permit various solutions to the problem of quantifying demand:

Stable Demand

For some markets, the supposition that demand is stable is a reasonable simplifying assumption. In effect, in this type of market, fundamental price forecasts can be based strictly on supply statistics.

Growth Pattern in Demand Change

For other markets, although demand changes from year to year, the pattern of change can be described by a simplified assumption (e.g., demand increases by 3 percent annually). For markets of this type, demand can be represented by an index that changes in a manner consistent with the assumed growth pattern for demand.

Identification of Demand-Influencing Variables

For some markets, although changes in demand cannot be described by any consistent growth pattern, the factors that affect demand can be identified. For example, beef demand increases in some years and decreases in others. Nevertheless, it can easily be demonstrated that these shifts are dependent on other identifiable factors, such as availability of competitive meat supplies. In such cases, one can bypass the problem of precisely specifying the demand curve by directly formulating a price-forecasting model that uses supply statistics and the factors determining demand as inputs. An example of such a model is given by the following equation:

$$QCP = f(CS, HS, BS, T)$$

where QCP = average quarterly cattle price
CS = quarterly cattle slaughter
HS = quarterly hog slaughter
BS = quarterly broiler slaughter
T = time trend

In the preceding example, CS represents a supply variable, while HS, BS, and T represent variables that affect demand. Trend affects demand through inflation (more will be demanded at each nominal price level because of inflation) and possible other factors that have a trending characteristic.

As another example, in attempting to forecast copper prices, one of the ways we could incorporate the demand effect would be by focusing on the level of activity in the key copper-using industries. Figure 22.11 illustrates the relationship between copper prices and an index of new housing for the same 21-year period that was surveyed in Figure 22.10. Figure 22.12 illustrates the relationship between copper prices and domestic auto sales during the same period. Note how the

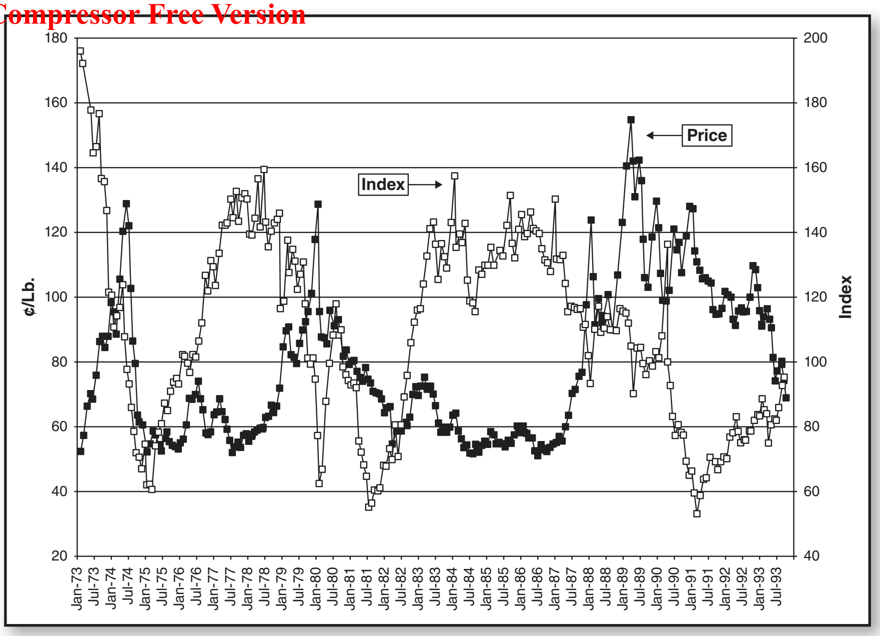

FIGURE 22.11 Average Monthly Copper Nearest Futures Price vs. Index of New Private Housing

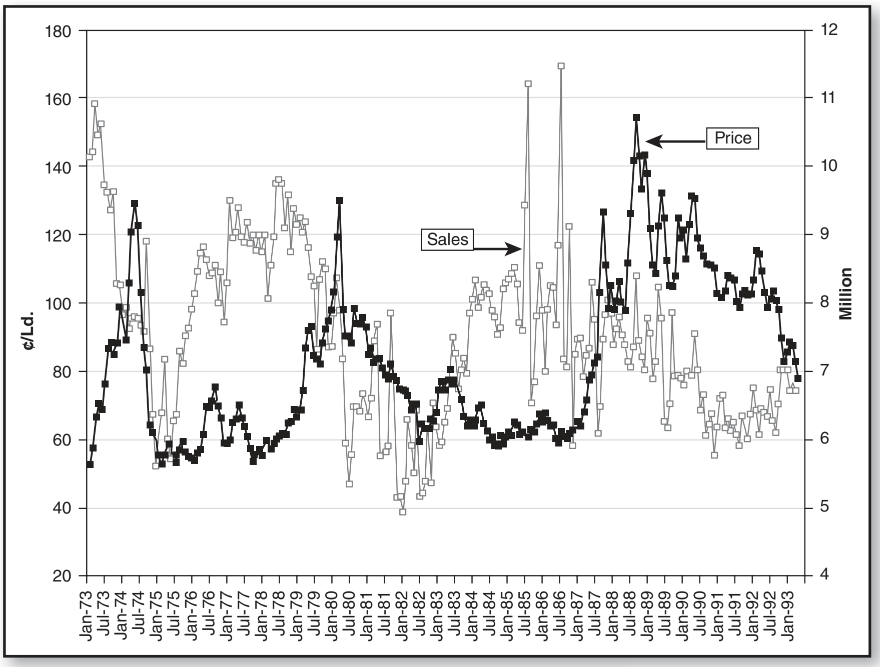

FIGURE 22.12 Average Monthly Copper Nearest Futures Price vs. Annualized Seasonally Adjusted Auto Sales

declines in housing starts and automobile sales *preceded* downturns in copper prices, including the 1980 to mid-1982 decline. Recall from the previous section that the imposing 1980 to mid-1982 bear market seemed somewhat puzzling when viewed solely relative to the stock/consumption ratio. Figure 22.11 and 22.12 illustrate how this seeming paradox can be resolved once demand factors are considered. Of course, the specific demand factors included would change over time. For example, our copper illustration focused on the 1973–1993 time segment. In a current copper price model, indicators of emerging market demand would be far more critical than they were then.

Highly Inelastic Demand (and Supply Elastic Relative to Demand)

Although conceptually incorrect, practically speaking, for markets of this type it is possible to use consumption as a proxy for demand. Since by definition in these markets consumption in a given year will not vary widely, regardless of price level, one can assume the prevailing consumption level roughly reflects the demand level. For example, Figure 22.13 illustrates a series of inelastic demand curves and two different supply curves. Note how the quantity consumed at the equilibrium price level is primarily dependent on the prevailing demand curve. Hence, consumption can serve as a proxy indicator for the unknown demand curve.

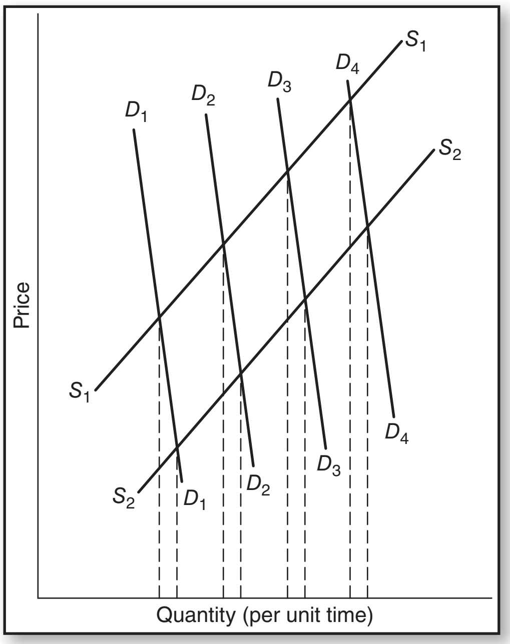

FIGURE 22.13 Consumption as Proxy for Inelastic Demand

An example of this approach is provided by the following model:

$$DASP = f\left(\frac{IS + P}{C}\right)$$

where $DASP$ = deflated average annual sugar price
IS = initial stocks
P = production
C = consumption

Note that initial stocks plus production is a proxy for supply, and consumption is a proxy for demand.

Why Traditional Fundamental Analysis Doesn't Work in the Gold Market

Unfortunately, the approaches we have enumerated for dealing with the elusiveness of demand do not encompass all cases. For some markets, not only is demand highly erratic, it is also difficult or nearly impossible to define a stable relationship that describes the precise dependence of demand on other variables.

Gold is a perfect example of such a market. Gold demand is basically dependent on the market's psychological perception of the value of gold, which in turn is dependent on a myriad of interrelated variables, including relative inflation rates, world interest rates, currency fluctuations, trade balance figures, OPEC actions, and political turmoil. The problem of specifying gold demand is further complicated by the fact that the relative importance of any of these factors in influencing gold demand is subject to considerable variation. For example, during some periods, currency fluctuations may become the pivotal price-influencing factor, while at other times developments in this area exert only a minor price impact.

In the case of gold, even the supply side of the equation cannot be readily approximated. Similar to demand, supply is subject to wide, erratic shifts that are also dependent on market psychology. This instability of the supply curve is primarily attributable to shifts in dishoarding rather than to changes in commercial supply.

The combination of highly erratic, intangible supply and demand curves makes the gold market a fundamental analyst's nightmare. Some analysts attempt to construct a fundamental model for gold by focusing on such statistics as mine production and industrial usage. This approach represents true folly, since these figures are equal to only a minuscule fraction of total gold supply. Gold prices are dependent on the psychological considerations detailed earlier, and there is no way to avoid this fact.

In effect, for a market such as gold, the traditional fundamental approach just does not work. Constructing an econometric model to predict gold prices is like trying to write a computer program that will predict a photographer's next picture on the basis of her past shots—the answer may be somewhat better than a blind guess but is hardly worth the effort.

Types of Fundamental Analysis

When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginnings of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.

—William Thomson, Lord Kelvin

The "Old Hand" Approach

The "old hand" approach refers to the analytical method used by analysts whose familiarity with the market is so finely honed that they have developed a virtual sixth sense with respect to its price fluctuations. By talking to a variety of commercial participants, they get a feel for market tone. They are also well tuned in to the flow of market news and are constantly assessing the market's behavior in response to this information. This is strictly a nonscientific approach, with the individual acting as the computer. It is not intrinsically inferior to more sophisticated approaches; its value is strictly dependent on the skills and intuition of the practitioner. In fact, it is hardly unusual for some analysts of this school to consistently outperform their econometrically oriented counterparts. This approach is strictly individualistic, however, and by definition can only be acquired by personal experience.

The Balance Table

The balance table summarizes the key components of current-season supply and disappearance, along with prior-season comparisons. The balance between supply and disappearance will indicate a season-ending carryover; it is the relative magnitude of this figure that is considered the primary price-determining statistic. Table 23.1 illustrates a U.S. Department of Agriculture (USDA) balance table

TABLE 23.1 U.S. Wheat Supply/Disappearance Balance, June–May Crop Year (million bushels)

	1988–1989	1989–1990	1990–1991	1991–1992	1992–1993	1993–1994[a]
Beginning stocks	1,261	702	536	366	472	529
Imports	23	23	37	41	70	75
Production	1,812	2,037	2,736	1,981	2,459	2,493
Total supply	**3,096**	**2,762**	**3,309**	**2,888**	**3,001**	**3,097**
Food use	726	749	786	789	829	845
Seed use	103	100	90	94	93	94
Feed/Residual	146	143	500	254	196	325
Total domestic use	975	992	1,376	1,137	1,118	1,264
Exports	1,419	1,233	1,068	1,280	1,354	1,125
Total disappearance	**2,394**	**2,225**	**2,444**	**2,416**	**2,472**	**2,389**
Ending stocks	**702**	**536**	**866**	**472**	**529**	**708**
Ending stocks as % of total use	29	24	35	20	21	30

[a]Projected.
Source: USDA.

for the wheat market. The analyst who relies heavily on the balance table will focus on possible shifts in the various components of supply and disappearance in an effort to anticipate the probable direction of price change.

The balance table is a valuable aid that succinctly summarizes the key market statistics. By itself, however, the balance table is insufficient in answering the critical question of what price is right under the given conditions. In fact, the analyst who uses only the balance-table approach to forecast prices will be guilty of fallacy number 1 detailed in Chapter 21 (i.e., viewing fundamentals in a vacuum).

The Analogous Season Method

In the *analogous season method,* the analyst finds past seasons that shared the same fundamental characteristics of a current season and then uses the price profiles of those analogous seasons as a "road map" in projecting current season price swings. For example, if in the current season production is up, usage is down, and the ending stocks/usage ratio is down, the analyst might find all past seasons that also exhibited these conditions. Next, the analyst would identify key price turning points in the analogous seasons (e.g., harvest low, postharvest high, winter low, crop-scare high). The timing and relative magnitude of price swings between these key turning points would then be calculated for each past analogous season. Finally, price swing ranges and turning point time windows would be projected for the current season, based on the assumption that the current season price patterns would be at least roughly similar to the price action of prior analogous seasons.

Regression Analysis

How can you determine which fundamental factors are most important in determining price levels? Even assuming you can make a reasonable conjecture as to what are the key fundamental factors influencing prices, how can you translate the current levels of these factors into a price forecast? Say, for example, you are trying to forecast hog futures prices. You assume that hog prices will be inversely correlated with hog slaughter levels and also inversely correlated with competitive meat supplies (e.g., broiler slaughter, cattle slaughter). You also assume that for any combination of supply levels for these various meats, prices will be higher in the current year than in past years because of the influence of inflation. Even if all these assumptions are correct, how can you determine the price implications for any given combination of the various meat supplies?

Simply comparing current supply levels to past year levels will not yield any price forecast. For example, what are the price implications of hog slaughter being 3 percent lower than in some prior year while broiler slaughter and cattle slaughter are each 2 percent higher? How does one reconcile the multiple comparisons of the current year to each of the past years examined? How much difference does a given time separation make when drawing comparisons between different years? All of these questions seem impossible to answer by simply comparing current and past data.

Regression analysis provides a statistical procedure that can be used to translate fundamental data into price projections. The assumptions we just made regarding the plausible key influences on hog prices could be formalized into the following equation:

$$P = a + b_1H + b_2B + b_3C + b_4T$$

where P = average price
H = hog slaughter
B = broiler slaughter
C = cattle slaughter
T = time trend

The values of a, b_1, b_2, b_3, and b_4 are determined by the regression analysis procedure (explained in the appendices). Given projections for hog slaughter, broiler slaughter, and cattle slaughter, we can plug those values and the time trend value for the current year into the preceding equation and obtain a precise price forecast.

Even if you are not mathematically inclined, think twice before dismissing the regression-analysis approach. Regression analysis embeds a number of important attributes:

1. Regression analysis makes it possible to combine multiple fundamental inputs, compared across multiple years, to derive a price forecast.
2. Regression analysis can be used to test the relative significance of each of the price-influencing variables (called *independent* variables) as well as the forecasting equation as a whole.
3. Regression analysis provides an efficient learning tool for understanding the interrelationships between various fundamental factors and price.

Regression analysis is probably the single most useful analytical tool in fundamental analysis. Appendices A through F provide an in-depth discussion of regression analysis.

Index Models

Sometimes we may wish to construct a fundamental model that uses scores of explanatory variables as indicators of a market's price. For example, we might postulate that bond prices are inversely related to a variety of inflation indicators (e.g., gold prices, S&P Goldman Sachs Commodity Index, consumer price index), economic indicators (e.g., employment, industrial production, housing starts), and monetary indicators (e.g., yield spread). Given the wide range of such indicators, and allowing that each indicator can be used with multiple time lags (as the relationship between bond rates and an indicator will frequently not be contemporaneous), it is easy to see how the number of possible explanatory variables could reach 50 or even higher.

Regression analysis cannot handle situations that involve large numbers of explanatory (independent) variables. Typically, a regression equation will employ five or fewer independent variables. There are two primary reasons why regression analysis cannot be applied to cases involving a multitude of variables:

1. If large numbers of independent variables are used, there is a great danger of overfitting (i.e., deriving a model that is tailored to fit past data but will be useless as a tool for projecting future prices or price trends).
2. When large numbers of independent variables are employed it is virtually inevitable that a number of such variables will be closely related to each other. High correlations among independent variables in a regression model will result in a statistical problem called *multicollinearity,* which destroys the reliability of the derived forecasting equation. (This problem is discussed in greater detail in Appendix E.)

One method of handling a large number of explanatory variables is to combine them all in an index model. The following step-by-step approach illustrates one possible procedure:

1. Assign each indicator a value of +1 if its current status is considered bullish for the price of the given market and a value of −1 if it is considered bearish. (How such a determination is made will be discussed momentarily.)
2. Add all the *assigned* indicator values to obtain an index value.
3. Normalize the index by multiplying by 100 divided by number of indicators. This step will yield an index with a theoretical range of −100 (if all the indicators are bearish) to +100 (if all the indicators are bullish). For example, if there are 50 indicators, and 30 are bullish and 20 bearish, the preceding procedure would yield a normalized index value of +20. Of course, an equal split between bullish and bearish indicators would yield an index value of 0, as is intuitively desirable.

This procedure sounds simple enough. The key question, however, is: how does one determine if a current indicator value is bullish or bearish? Deciding on some value as the division line between bullish and bearish values is highly undesirable for two key reasons: (1) many variables trend over time; and (2) due to structural changes over time, the definition of "high" and "low" will tend to shift for many, if not most, variables. Hence, it is far more practical to categorize a variable as bullish or bearish based on its *direction* of movement (i.e., trend) rather than its *level*. Trend categorization, however, falls more within the realm of technical analysis than fundamental analysis. Indeed, some of the very basic tools of technical analysis (e.g., crossover moving average) can be applied to defining the assigned indicator values in an index model of the type described in this section. For example, if using a crossover moving average to define the trend direction of the indicators, an indicator would be assigned a value of +1 if the short-term moving average was greater than the long-term moving average, and a value of −1 in the reverse case.

CHAPTER 24

The Role of Expectations

What we anticipate seldom occurs; what we least expect generally happens.

—Benjamin Disraeli

Using Prior-Year Estimates Rather Than Revised Statistics

Historical data are based on final revised estimates rather than the estimates that were available at the time. For example, the historical levels of U.S. corn production are revised throughout the season with the final revision occurring after the end of the season. These final revised estimates for each season (the *actual* levels) can differ substantially from the crop estimates that prevailed during each season (the *expected* levels). Similarly, historical corn consumption and export levels (the actual levels based on final revised estimates) can be very different from the expected levels that prevailed during each season.

Typically, fundamental models would use actual historical data as inputs. But is this default approach the best procedure? A strong argument can be made that the data levels expected at the time are more relevant to explaining price behavior than actual data levels that only became known after the price forecast period in question. Thus, it may be possible to build a more accurate model using past estimates rather than actual statistics as the price-explanatory variables. For example, if we are trying to construct a model to explain and predict September–November corn prices, we might well find that the past production and usage estimates released during the September–November period are more helpful than the actual supply statistics in explaining the year-to-year historical variation in September–November prices. Such price behavior would merely reflect that what the market thought was true in the past was more important in determining prices than what was actually true (as defined by the final revised estimates)—a reasonable outcome given that market participants have no way of determining actual statistics and must rely on prevailing estimates for their marketing, purchasing, and trading decisions.

The key point is that using past expected data rather than actual statistics might be theoretically sounder and may well yield better price-forecasting models. Of course, using past expected statistics in a model will require considerably more work in terms of data gathering, which may explain why such data are far less frequently used than the final revised numbers. As is true for most of life's endeavors, creating a better product (price-forecasting model in this case) requires more effort. There are no shortcuts in doing things right.

Adding Expectations as a Variable in the Price-Forecasting Model

Thus far, we have discussed the choice between using anticipated versus actual data for explaining past price variation. Regardless of which is used, one can also consider adding a variable to represent expectations for a key statistic in a following season. To clarify this distinction, we list the four possible variations in the extent to which expectations are incorporated in a model:

1. Price is a function of concurrent-season actual statistics (no use of expectations).
2. Price is a function of estimates for concurrent-season statistics (expectations used for concurrent-season data).
3. Price is a function of concurrent-season actual statistics and expectations for the following season (expectations used for following-season data).
4. Price is a function of concurrent-season estimates and expectations for the following season (expectations used for both concurrent- and following-season data).

The expectations for a coming season can often exert a more pronounced price impact than do prevailing fundamentals. This observation is especially true during the latter half of a season—a time at which the fundamentals for the given season are usually well defined and not subject to large variation. In fact, frequently, when there is a dichotomy between the implications of old-crop fundamentals and new-crop expectations, the latter tend to dominate the price picture.

Why should expectations for a coming season affect current-season prices? Expectations influence current selling and buying psychology. For example, if supplies are burdensome and a shift toward supply tightness is anticipated, sellers will have an incentive to hold back the commodity and will offer less to the market at each given price level (the supply curve will shift upward in response to reduced supply). At the same time, buyers will attempt to build inventories and therefore will purchase increased quantities at any given price level (the demand curve will shift upward). These two effects reinforce each other, and the net result will be higher prices in the current season.

The Influence of Expectations on Actual Statistics

Ironically, bullish new-crop expectations can actually cause current-season fundamentals to appear more bearish. The following cause-and-effect diagram illustrates this point:

Bullish expectations for new crop → price during old-crop season ↑ →
old-crop consumption and exports ↓ → old-crop stocks ↑

As a result of this string of events, seasons that experience bullish new-crop expectations are likely to appear inexplicably overpriced based on old-crop fundamentals—another reason why new-crop expectations should be incorporated in the price-forecasting model wherever possible.

Defining New-Crop Expectations

On the supply side, new-crop expectations can be based on planting intentions and subsequently on acreage estimates. In using these estimates to define expectations, one usually assumes a trend yield (the yield implied by a regression-derived best-fit line of past yields) or an average yield (e.g., five-year average for each state or region) if there is no pronounced trend. Such neutral projections would then be adjusted upward in the case of very favorable growing weather, or downward if conditions were adverse.

On the usage side, expectations are defined by the historical behavior pattern. For example, if in recent years consumption changes for a given commodity have tended to range from −2 percent to +4 percent, as a function of the direction and magnitude of price change, in the absence of any additional information, one might use a 1 percent consumption increase as a representative figure for expected new-crop consumption.

Historical expectation statistics can be generated in a similar manner or by surveying past commentaries in U.S. Department of Agriculture situation reports, trade reports, and industry market reports. Unfortunately, there is some unavoidable arbitrariness in the latter approach, since the expectation figures depend on the sources chosen and on the weights assigned to each source. However, this ambiguity is not a critical drawback, since at any given time, new-crop projections by various sources tend to cluster in the same general area.