CHAPTER 8

Money Management

Everybody's got a plan until they get hit.

—Mike Tyson

Although it is not true that a good money management system can turn any trading method into a profitable one, it is certainly possible to ruin even the best trades with poor money management. Edge estimation and capture are difficult and both involve subjective judgment. So it is probably understandable that traders focus so heavily on these aspects of trading. But money management and trade sizing is just as essential for success. In this chapter we look at the various methods we can use to size trades and the risk/reward characteristics that each choice implies.

Ad Hoc Sizing Schemes

Consider the equity curves shown in Figures 8.1 and 8.2. It seems obvious that the trader in the first figure is in some sense better. He has made much more money in the same period of time, albeit while displaying more variance. Actually, in this instance, the trades that the traders made were identical. All the difference in their equity curve was due to the sizing strategy. Both traders were playing a game where they flipped a coin and received a dollar if they were correct and paid a dollar if they were wrong. Each won 550 bets out of 1,000. But the first trader finished with \$5,207 and the second only accumulated \$620. Trade sizing is clearly an important issue.

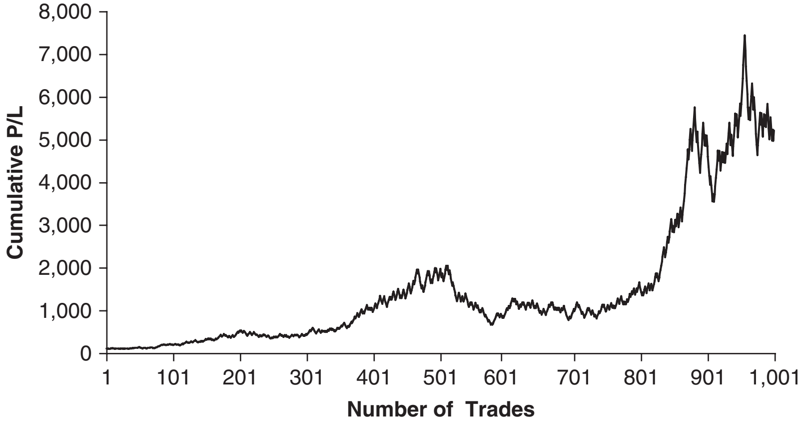

FIGURE 8.1 An Equity Curve for a Trader Betting 5 Percent of His Bankroll on Each Trade

FIGURE 8.2 An Equity Curve for a Trader Betting $5 on Each Trade

The simplest method is to size according to "feel." The trade size is adjusted on an ad hoc basis depending on how good the trader thinks the individual bet is. This is a terrible idea. This isn't really so different to picking trades based on feel and it is quite possible that the sizing decisions can dominate the trade choices themselves. Our entire methodology is based on the idea that we can systematically approach the trading process, and to size our trades according to hunches goes completely against this. It will let us fall victim to our moods and psychological biases: precisely what we are trying to avoid.

The next step is to trade the same amount every time. For example, we might choose to always trade 100 options or 1,000 vega. This is known as a *fixed trade size system*. Any system that can find value will be profitable (over the long run) with such a sizing scheme. In back testing this is often the sizing scheme used, as it most clearly shows the profitability of the underlying method.

Or we could choose to trade a constant percentage of our bankroll at each opportunity. This is known as a *fixed fraction sizing system*.

Our example earlier was for just one group of 1,000 bets. Figure 8.3 shows the results for another three trials. We can see two things. First, even though in all cases we have the same mathematical edge (a 55 percent chance of winning an even money bet) the results vary wildly. Second, the choice of bet size makes a huge difference. Proportional betting sometimes results in vastly greater final wealth and our wealth can never go to zero (at least in the idealized case we consider here where we have infinitely divisible currency units), but looks to be much more volatile in its results than fixed betting.

This numerical experiment has clearly shown that the choice of a staking plan is important. And in addition to the two schemes mentioned earlier, there are a number of others we could dream up. For example, we could bet so that all successful bets would win the same amount, increase bets (either by absolute size or relative size) after a win, or we could adopt a similar strategy where we increase bet size after a loss. With all these schemes there will be parameters we need to choose, for example, what percentage of our bankroll to start betting with.

These methods are often used as the basis of betting systems. It is quite easy to make them seem appealing to mathematically unsophisticated gamblers. Let's consider again our coin-tossing game where we win at a rate of 55 percent. If we double our stake each time we lose a bet until we eventually win, we end many betting runs with wins of a single unit but eventually we lose our entire bankroll. We haven't changed the expectation of our distribution, but we have introduced a massive amount of skew, trading frequency of win for size of loss.

How can we decide between all the alternatives? A possible solution to this problem was found by John Kelly (1956).

The Kelly Criterion

Consider a general situation such that when we win, we gain w percent and when we lose, we lose l percent. Our bankroll is initially W_0. Each bet is a set fraction, f, of the bankroll.

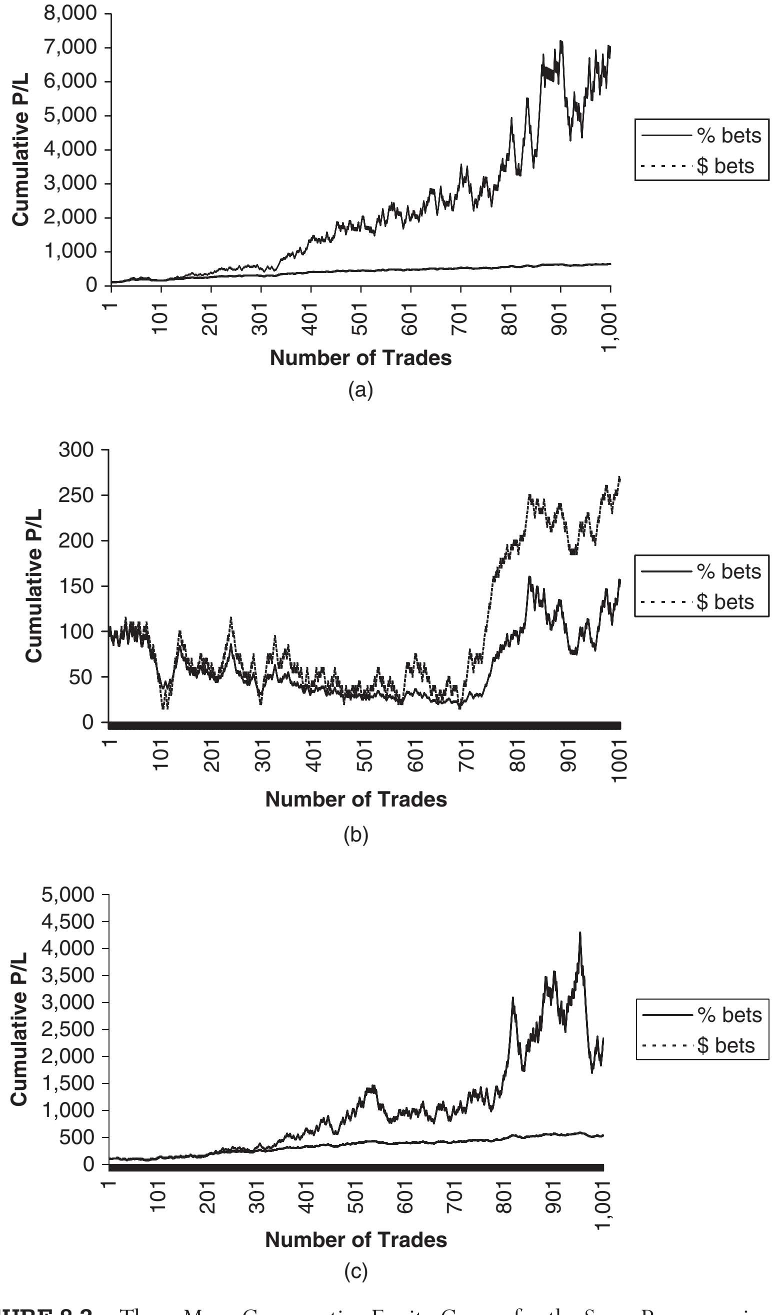

FIGURE 8.3 Three More Comparative Equity Curves for the Same Process as in Figure 8.1

So after a win our bankroll is

$$W_0(1 + fw) \tag{8.1}$$

Or we can say that the *gain factor* is $(1 + fw)$. Similarly, a loss would leave us with

$$W_0(1 - fl) \tag{8.2}$$

So the gain factor here is $(1 - fl)$. More simply, each time we win we multiply by $(1 + fw)$ and when we lose we multiply by $(1 - fl)$. So for n wins and m losses the gain factor is

$$G(f) = (1 + fw)^n(1 - fl)^m \tag{8.3}$$

Or per trade we get

$$G(f)^{\frac{1}{n+m}} \equiv g(f) = (1 + fw)^p(1 - fl)^q \tag{8.4}$$

where $p = n/(n + m)$ is the probability of a win and $q = m/(n + m)$ is the probability of a loss.

It would be a mistake to choose f to maximize G. On any finite number of bets, our expected return would be maximized by betting our entire bankroll each time. Sadly, this also gives us a 100 percent chance of going bankrupt, as eventually we are bound to lose. This strategy takes no account of risk. We actually want to maximize our risk-adjusted return or utility. This involves choosing a particular utility function again. Generally the log utility function is chosen (we could choose any one of a number of utility functions but the log function can be shown to do better in the long run than any other). In this case we will not be maximizing *expected* wealth but we will be maximizing *typical* wealth. This distinction is important. The average value would be heavily distorted by the unlikely case where the trader wins every trade. This won't be of great consolation to the traders who have gone bankrupt. By maximizing the logarithm of expected wealth we eliminate the possibility of bankruptcy.

So we take the log of the gain function and find the optimal f by differentiating with respect to f then setting this equal to zero. This gives

$$f = \frac{(pw - ql)}{wl} \tag{8.5}$$

(So in our initial simple example the Kelly fraction would have been 0.1.)

Our expected bankroll after N bets would be given by

$$W = W_0(1 + p\ln(1 + fw) + q\ln(1 - fl))^N \tag{8.6}$$

Figure 8.4 shows the expected bankroll for 10 bets as a function of bet size where wins pay twice as much as the losses and we win 45 percent of

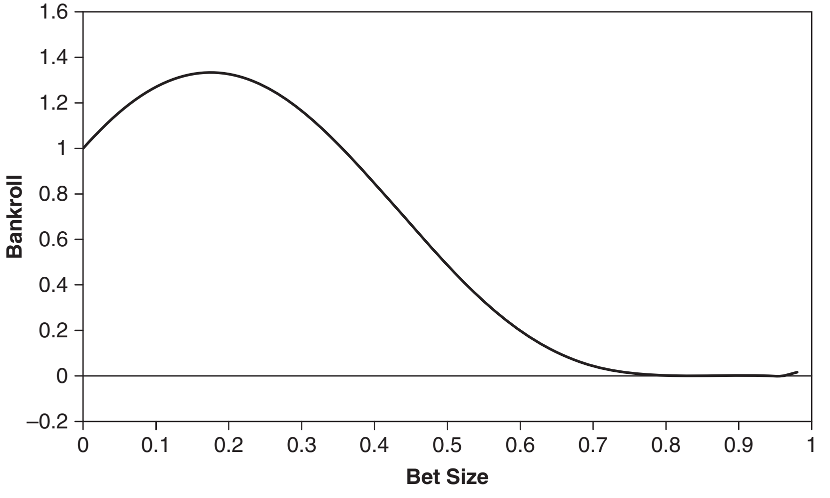

FIGURE 8.4 The Expected Bankroll for 10 Bets as a Function of Bet Size

bets. The Kelly ratio in this case is 0.175, which corresponds to the peak of the curve. Betting more than twice Kelly turns the growth rate negative and our bankroll becomes more depleted as we bet bigger.

Running some simple simulations should make several points abundantly clear:

- The Kelly fraction generally finishes with more wealth than any other chosen proportional scheme.
- The swings in our equity become uncomfortably large when betting the Kelly fraction.
- Betting more than Kelly is worse than betting less.

This last point is worth emphasizing. Betting more than Kelly results in higher volatility and lower returns. This can be seen in Figure 8.5 where we show one realization of a profit and loss (P/L) curve for the coin tossing game when using half Kelly, full Kelly, and twice Kelly bet sizes.

At this point, let's just assume that the Kelly strategy is sufficiently intriguing to make it worthwhile considering a more realistic situation: one that is closer to those we might encounter when trading financial instruments.

We need to generalize the situation to deal with a continuous outcome. Imagine trading a situation where the outcome of a bet or trade is known to have a certain distribution. This would be the typical situation facing a trader: Either the distribution can be estimated from historical trade results

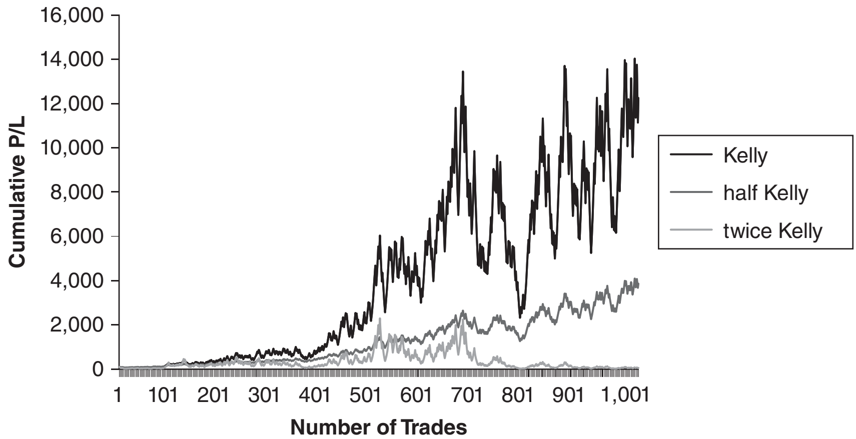

FIGURE 8.5 The Relative Volatility of a P/L Path Generated by Trading at Various Multiples of the Kelly Ratio

or from more theoretical considerations. We do still insist at this point that the trades be independent and identically distributed. Again we bet a fraction, f, of our wealth at the start of each period so that

$$W_n = W_{n-1} + f W_{n-1} g(X_n) \tag{8.7}$$

where X_n is the random variable giving the result of the nth trade and it has the payoff $g(X_n)$. After a sequence of n trades our bankroll will be

$$W_n = W_0 \prod_{i=1}^{n} (1 + fg(X_i)) \tag{8.8}$$

Now we take logarithms as before:

$$\ln(W_n) = W_0 \sum_{i=1}^{n} \ln(1 + fg(X_i)) \tag{8.9}$$

So

$$E[\ln(W_n)] = nW_0 E[\ln(1 + fg(X_n))] \tag{8.10}$$

$$= nW_0 \int \ln(1 + fg(x))\Phi(x)dx \tag{8.11}$$

where $\Phi(x)$ is the distribution function that describes the results of the trades. If we maximize over the bankroll fraction, f, we find that the optimal value is the one that satisfies

$$\int \frac{g(x)\Phi(x)dx}{1 + fg(x)} = E\left(\frac{g\,(x)}{1 + fg(x)}\right) = 0 \tag{8.12}$$

This equation is quite general. It applies to all distributions. It is sometimes incorrectly stated that the Kelly criterion only applies to trades with binary outcomes. This is untrue. However, in the case of binary events, certain simplifications can be made. These approximations do indeed have limited applicability. In general, Equation 8.12 is too unwieldy to use directly, but if we consider the case where our edge per trade is small we can make things simpler (and sadly this case usually is fairly representative of reality). In this situation we know that f will be small and we can expand Equation 8.12 in a power series, and then truncate after leading order to get

$$\begin{aligned} 0 &\approx \int g(x)\Phi(x)(1 - fg(x) + \cdots)dx \\ &= \int g(x)\Phi(x)dx - f\int g^2(x)\Phi(x)dx + \cdots \end{aligned} \tag{8.13}$$

But the first term is just the expected payoff for a unit bet and the second term is the variance of the payoff, $g(x)$. So in the limit of small edge we obtain

$$f = \frac{r}{\sigma^2} \tag{8.14}$$

This is certainly simple enough to use. To estimate our trading size we need only the expected return of the trade and its variance, no matter how complicated the actual trade is. Actually, the return here should generally be interpreted as the return over the risk-free rate. Many derivations neglect to mention this (those aimed at gamblers probably neglect it as it is difficult to earn interest while sitting at a blackjack table and bookmakers tend not to pay interest, either).

The expected growth rate for someone trading at a fraction, f, of the Kelly ratio is given by

$$GR = \left(f - \frac{f^2}{2}\right)\frac{r^2}{\sigma^2} \tag{8.15}$$

which is maximized for $f = 1$: trading at the full Kelly ratio when the growth rate is

$$GR_{\max} = \frac{r^2}{2\sigma^2} \tag{8.16}$$

And we see that the growth rate is zero for $f = 0$, the case where we don't trade at all, and $f = 2$, when we drastically overbet. As the growth rate is symmetric around $f = 1$, we can see it is better to bet more conservatively and hence to underestimate our edge (or equivalently to overestimate our variance) as we will obtain the same growth for $f = 1 - x$ and $f = 1 + x$.

If we are trading according to the Kelly criterion the probability that we reach a bankroll $B \times W > W_0$ before we dip to $A \times W < W_0$ is given by

$$P(A,B) = \frac{1 - A^{1-\frac{2}{f}}}{B^{1-\frac{2}{f}} - A^{1-\frac{2}{f}}} \tag{8.17}$$

It is interesting that the edge and variance of the trade don't appear here. Having a "better" trade just speeds up the process. If $A = 0.5$ and $B = 2$, we get $P(A,B) = 2/3$. Conversely, this means that when betting at full Kelly you have a one-third chance of having your bankroll halve before it doubles! As we saw in our earlier simulations Kelly sizing is exceedingly volatile. To deal with these extreme drawdowns it is reasonably common for Kelly devotees to trade at a fraction of the Kelly ratio. Table 8.1 shows how the probability of halving before doubling can be changed by using a fractional Kelly ratio.

Reducing drawdowns by trading at a fraction of the Kelly ratio is not a free lunch. By trading smaller we dramatically increase the time needed to reach our upside goal. The expected exit time is given by

$$E[T] = \frac{1}{GR} \log\left(\frac{B^{P(A,B)}}{A^{P(A,B)-1}}\right) \tag{8.18}$$

This is the mean time before we reach our goal $(B.W_0)$, or are stopped out (at $A.W_0$).

There is no compelling *theoretical* reason for sizing trades according to the fractional Kelly idea. Fractional Kelly doesn't correspond to maximizing any utility function. There are two *practical* reasons for using fractional Kelly.

1. It is a pragmatic attempt to find a middle ground between realizing the potential returns from trading at the full Kelly ratio and reducing the resulting high volatility.

TABLE 8.1 **Probability of Doubling before Halving as a Function of the Kelly Fraction That We Are Trading**

Fraction of Kelly	Prob (A,B)
1	0.667
0.8	0.739
0.6	0.834
0.4	0.941
0.2	0.998

2. It is a poor man's Bayesian approach to acknowledging the general difficulty of beating markets. For example, half-Kelly corresponds to taking an average of our edge estimate and the zero-edge alternative that the market is a perfect predictor of future value.

We now have some sort of intuition about our prospective profit and loss distribution when using Kelly. We have the expected growth rate, a measure of dispersion (the drawdown probability) and the expected time to reach our goals. But it is actually possible to calculate the entire probability distribution of our future bankroll. A paper by Chapman (2007) shows how the bankroll spreads through time. For convenience we take the initial bankroll to be one. He shows the probability distribution evolves as

$$P(x,f,t) = \exp\left(\left(1-\frac{1}{f}\right)t\right)\frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{\left(\log x+\left(\frac{3}{2}-\frac{1}{f}\right)t\right)^2}{2t}\right) \tag{8.19}$$

Figure 8.6 is worth examining carefully. It shows the evolution through time of the probability distribution function (PDF) of the bankroll when trading at the Kelly fraction. We see that over time the bankroll relaxes and diffuses away from its initial value. We can also see that when betting at the full Kelly ratio, our distribution of outcomes is highly skewed (this should have already been evident from the earlier discussions of drawdowns but some more visual evidence can emphasize the point). The peak of the probability distribution function is less than one. We know that in the long run the Kelly strategy comprehensively dominates other strategies but we also know the result of any one series of trades is volatile so may be poor in the short run (before the growth rate has had a chance to overwhelm the volatility) and the "long run" may take quite some time to arrive.

Figure 8.7 shows how trading at a reduced Kelly fraction (in this case half-Kelly) substantially shifts the peak of the PDF to the right while still maintaining the skewness, which makes large wins possible.

Conversely, Figure 8.8 shows that trading higher than Kelly (twice Kelly in this instance) means that the PDF is pulled toward zero as time passes.

Sizing bets/trades according to the Kelly criterion is a controversial topic (see Poundstone 2005 for a very readable account). Most of the discussion centers around the idea that maximizing the logarithm of expected wealth isn't really what a sensible investor would want; that is, the utility function is specified incorrectly. The anti-Kelly group includes Nobel Prize winners

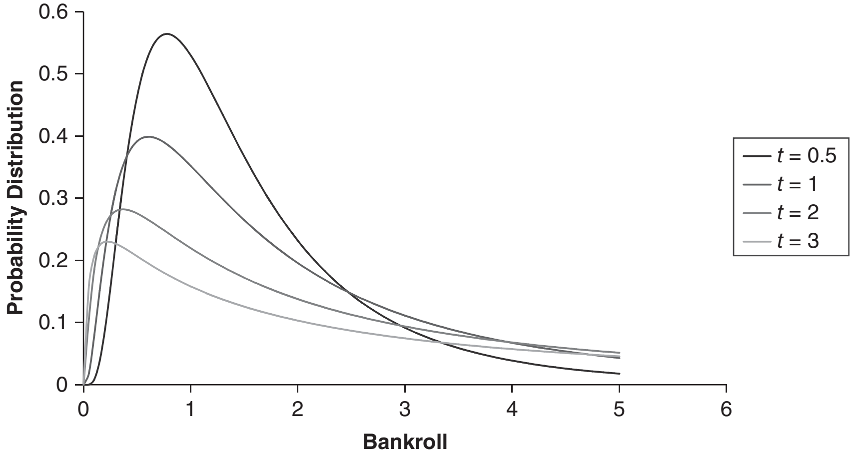

FIGURE 8.6 The Evolution through Time of the PDF of the Bankroll When Trading at the Kelly Fraction

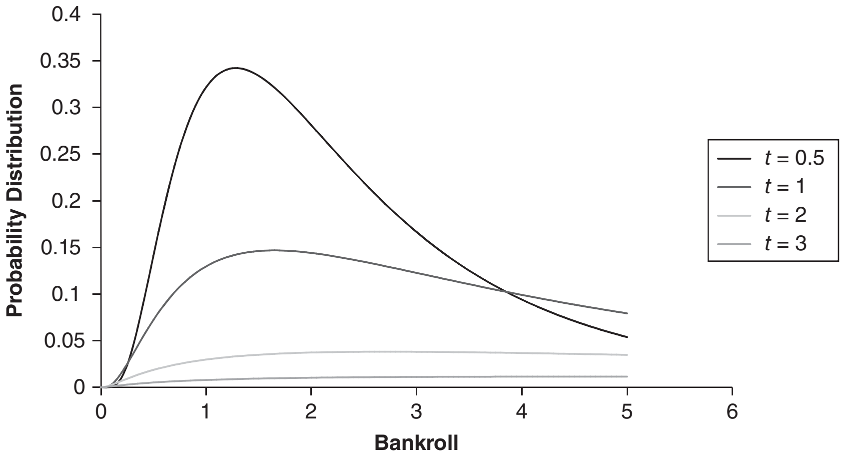

FIGURE 8.7 The Evolution through Time of the PDF of the Bankroll When Trading at the Half of the Kelly Fraction

(Samuelson 1979), finance professionals (Brown 2002), and sports gamblers (Miller, www.professionalgambler.com). However, the other side of the argument also boasts some impressive names, including Ed Thorpe (1984, 1997), Claude Shannon (inventor of information theory), David Shaw (founder of D. E. Shaw), and William Miller (manager of the Legg Mason Value Trust, the only SEC regulated mutual fund to outperform the S&P 500 for 10 consecutive years). As the method is dependent on a utility

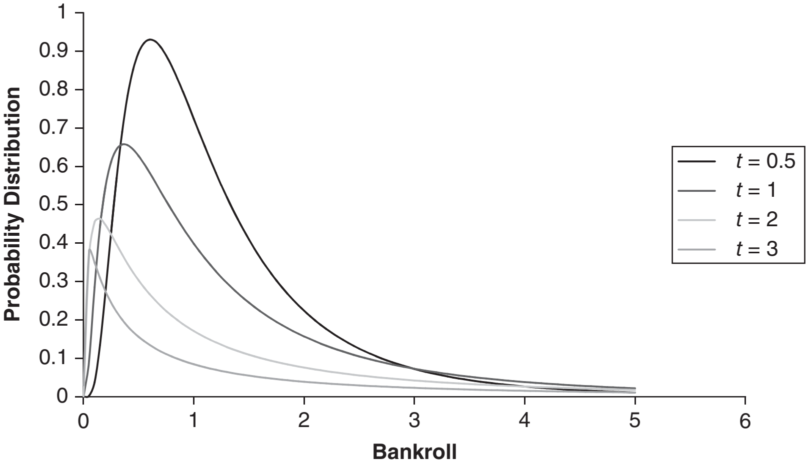

FIGURE 8.8 The Evolution through Time of the PDF of the Bankroll When Trading at Twice of the Kelly Fraction

function, it isn't too surprising that people disagree. But instead of getting into arguments over what other people's risk preferences should be, let's simply look at the Kelly criterion's good and bad points.

Good Points

- Maximizing the expected value of the logarithm of wealth asymptotically maximizes the rate of growth, so the Kelly strategy eventually outperforms all others.
- The Kelly strategy has zero risk of ruin.
- On average we will always be ahead of any other strategy.
- The strategy is myopic, in the sense that we only ever need to consider our current opportunities and bankroll, not subsequent situations. This is not the case for progression-type methodologies where current trade sizes are a function of previous trade sizes. Myopia is useful when deciding if a strategy is practically useable.
- Trading a fraction of the Kelly amount allows us to easily tune our desired level of risk at the expense of lower expected returns.

Bad Points

- When betting a fraction of wealth, a loss followed by a win still leaves us behind.

- The amount bet becomes extremely large in situations where we have a large amount of edge, that is, when the probability of a win is high or the risk is very low.
- Probability estimation becomes crucial. Overinvesting based on overestimation of success likelihood will lead to disaster.
- The total amount of money invested is far larger than the winnings.
- Due to the strategy's volatile outcomes, it is possible to have sessions with poor outcomes even though the long-term expectations are high.
- The time necessary for the long run to dominate can be very long indeed.
- Sometimes it isn't obvious what "bankroll" is.

The good points have been studied and may well seem so good that any bad points can be ignored. Isn't the fact that sizing according to the Kelly criterion eventually creates more wealth than any other strategy, enough to override any concerns? Before we come to this conclusion, let's take a slightly closer look at the bad points we haven't already noted.

Time for Kelly to Dominate

We all know we need to be patient when trading. This is often axiomatically stated as a virtue. We tend to say reflexively that we possess patience because we are so acutely aware that it is necessary for a good trader, but how patient do we need to be to reap the benefits of the Kelly strategy?

An example from Browne (2000) shows that waiting for the long run can be more tedious than we might reasonably expect. He considers an investor who has a choice between a stock with an annual return of 15 percent and a volatility of 30 percent and an interest bearing account that pays 7 percent. The Kelly criterion (Equation 8.14) has us invest 89 percent of our wealth ($0.15 - 0.07/(0.3)^2$) in the stock and put the remainder in the bank. Before we have a 95 percent probability of beating the all cash portfolio by 10 percent, we have to wait 157 years. Even worse, to have a 95 percent probability of beating the all-stock portfolio by 10 percent, we have to wait 10,286 years. Even the expected times to outperform these benchmarks is 2.8 years and 184 years.[1] Patience is indeed necessary.

[1]The expected time for Kelly to outperform another strategy with a trading fraction f', by ε percent is given by $\frac{2}{\sigma^2(f-f')^2}\ln(1+\varepsilon)$

Effect of Parameter Mis-Estimation

In practical financial applications we will never really know the distribution of outcomes of the trades. Do our errors in estimating the distribution actually matter? This was studied by Medo, Pis'mak, and Zhang (2008).

First we consider trades with a binary outcome. If we lose, we pay a dollar and if we win, we gain a dollar. If the win probability is $p > 0.5$, we have a favorable game and the Kelly ratio is

$$f = 2p - 1 \tag{8.20}$$

But what if we don't *know* p? What if we have to estimate it from historical data? Assume that we have N trials to do this estimation and in this period we see w wins. Bayes' theorem gives the distribution of the true probability as

$$\theta(p|w, N) = \frac{\pi(p)P(w|p, N)}{\int_0^1 \pi(p)P(w|p, N)dp} \tag{8.21}$$

where $\pi(p)$ is the prior distribution of p, and $P(w|p,N)$ is the probability distribution of w, given N and p. P is a binomial distribution with mean pN.

$$P(w|p, N) = \binom{N}{w} p^w (1 - p)^{N-w} \tag{8.22}$$

As *all* of the information we have is contained in the observations, the prior needs to reflect this maximum ignorance. So we need to use the uniform distribution, $\pi(p) = 1$, over the range of p from zero to one (The choice or prior can also be used to bias our estimate low, reflecting caution in our ability. For example, we might choose to use a distribution defined over the range of zero to 0.75). Evaluating the integral in Equation 8.21 gives us

$$\theta(p|w, N) = \frac{(N + 1)!}{w!(N - w)!} p^w (1 - p)^{N-w} \tag{8.23}$$

This distribution has a mean

$$\langle p \rangle = \frac{w + 1}{N + 2} \tag{8.24}$$

This is important. Our naive guess of win rate (6 wins out of 10 implying a true rate of 0.6) will always overestimate. We need to incorporate the effects of our prior ignorance. This lessens the weight we give to the observations. The distribution is shown in Figure 8.9 for $w = 6$ and $N = 10$. Note also that there is a significant chance of the results of this trade

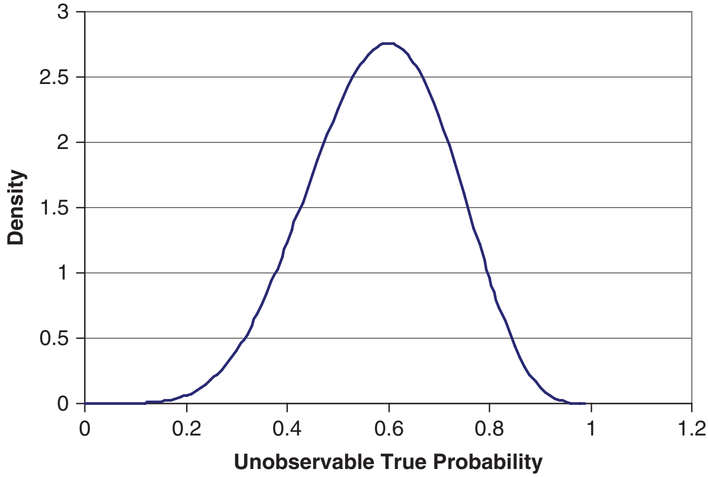

FIGURE 8.9 The Posterior Distribution When Observing 6 Wins in 10 Plays

appearing to have a negative expectation. This can be found by looking at how much of the cumulative Kelly distribution is below 0.5. In this case it is around 27 percent.

Further, we can calculate the variance of the distribution.

$$\mathrm{Var}(p) = \frac{(w+1)(w+2)}{(N+2)(N+3)} - \frac{(w+1)^2}{(N+2)^2} \tag{8.25}$$

In this example, we get a standard deviation of 0.137.

What is the Kelly ratio in these circumstances, where clearly we have the extra "risk" of parameter uncertainty? There is a significant chance that we may even be playing in a losing game.

The Kelly ratio is the amount, f, that maximizes the logarithmic gain. We also know, from the work of Chapman (among others), that this corresponds to using the average of the win probability. But we can illustrate this with a simple example.

Consider the simplest case. The winning probability can take one of two values: p_1 with probability F_1 or p_2 with probability F_2. Now we can write the growth rate as

$$G = (F_1 p_1 + F_2 p_2)\ln(1+f) + (1 - F_1 p_1 - F_2 p_2)\ln(1-f) \tag{8.26}$$

Differentiating this with respect to f and setting equal to zero gives the Kelly ratio as

$$f = 2(F_1 p_1 + F_2 p_2) - 1 \tag{8.27}$$

This is clearly just

$$f = 2\langle p \rangle - 1 \tag{8.28}$$

This doesn't seem particularly interesting, but when we combine this with Equation 8.24 we see that

$$f = \frac{2w - N}{N + 2} \tag{8.29}$$

If we compare this with the naive value where we estimate p to be w/N we see that

$$\frac{f}{f_{naive}} = \frac{\dfrac{2w - N}{N + 2}}{2\dfrac{w}{N} - 1} = \frac{2w - N}{N + 2} \times \frac{N}{2w - N} = \frac{N}{N + 2} \tag{8.30}$$

which implies that the naive estimate will always bias our bets too high.

This effect isn't particularly large and in the limit of large N it disappears completely. In fact after 100 trials we will be off by only 2 percent. Of more importance is the variance of f. How many trials do we need to be reasonably sure that our measured value is close to being correct?
The delta method tells us that the variance of f is given by

$$\mathrm{var}(f) = \left(\frac{\partial f}{\partial p}\right)^2 \mathrm{var}(p) \tag{8.31}$$

Using Equations 8.25 and 8.31 we get

$$\mathrm{var}(f) = 4\left(\frac{(w + 1)(w + 2)}{(N + 2)(N + 3)} - \frac{(w + 1)^2}{(N + 2)^2}\right) \tag{8.32}$$

So looking back at the case in the figure where $w = 6$, $N = 10$: We would estimate that the standard deviation of f would be approximately 0.274. This again diminishes as N increases. If we had observed 60 wins in 100 trials our standard deviation would be only 0.097.

The actual dependence of the standard deviation of the Kelly ratio with sample size conforms to our expectations. More data implies less deviation. For a theoretical win probability of 0.6, the deviation as a function of sample size is shown in Figure 8.10.

What Is Bankroll?

Bankroll certainly isn't haircut. The haircut is the amount we need to post at our clearing firm, but it certainly is not the amount you are able to lose. Generally it is better to think of bankroll as the amount you can lose before the strategy is abandoned. But sometimes this isn't clear either. This

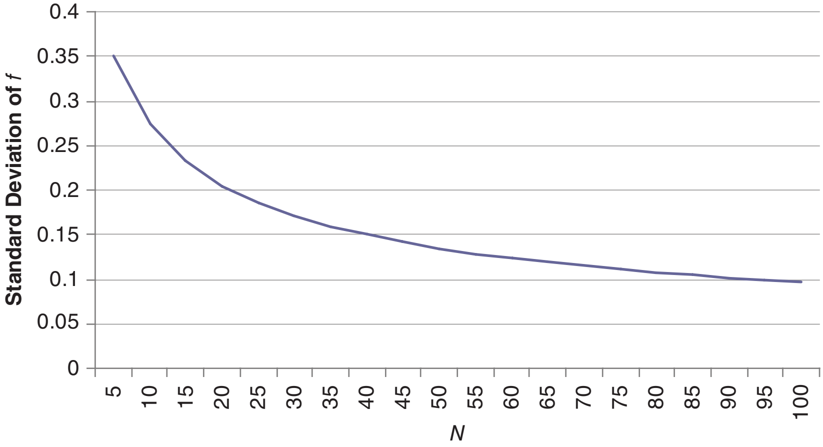

FIGURE 8.10 The Dependence of the Standard Deviation of f as a Function of Sample Size

was studied by Leib (1995). He pointed out that amateur blackjack players should bet much more aggressively than professionals as the amateurs could replenish their bankrolls from their other income. This concept also applies to traders who can switch jobs.

There are two ways that the Kelly criterion can be interpreted:

1. It is the strategy that maximizes the expect growth of the bankroll.
2. It is the strategy that maximizes the logarithmic utility function.

Neither interpretation is really consistent with a situation where we can replenish our bankroll.

Within the first framework, the goal is to become as rich as possible as quickly as possible. Going bankrupt is bad because it makes further growth impossible. Having any money at all left is infinitely better than having none. However, if there is another source of capital available (from another job or backer, for example), then further growth *is* still possible.

Under the utility growth interpretation capital growth is just a consequence. Utility growth is an end in itself. Here, going broke is infinitely bad purely because utility becomes unboundedly bad.

Under either interpretation, losing all of your bankroll is infinitely bad. It is important to emphasize the word *infinitely*. This means that there is never any way to come back. And this is seldom the case in real life. We are really talking about cases where the gambler has risked life itself.

We can align ourselves more closely with the Kelly approach if we consider our total wealth and then adjust the fractional adjustment accordingly.

For example, if we have a trading account of \$1 million and a total net worth of \$5 million we could choose a bankroll of \$5 million and a Kelly fractional multiplier of 0.1 instead of a bankroll of \$1 million and a multiplier of 0.5. For bets with small edge the differences will be minimal but when the edge becomes large the difference will be apparent.

Alternatives to Kelly

So is what Kelly does really what a trader wants? As always, the answer is, "it depends." What externally imposed constraints does he face? In a personal account where no one else can stop him out due to a drawdown, a trader may be happy to use Kelly and tune the volatility by trading a fraction of the full Kelly ratio. Similarly, when we are making an enormous number of trades in a short period we may accept a Kelly-based sizing method, as we can be more confident that the long run will arrive soon enough to overwhelm any variance effects. But when trading in an institutional setting or when being backed by someone else, the Kelly criterion is probably not really aligned with the interest of the trader. Here traders aren't as interested in optimal long-term growth as they are in a better chance of making a profit. They will trade maximal long-term potential for more certain short-term profits. We said earlier that Kelly was expected to outperform other strategies, so what can traders do that are better? They can gain a more certain short-term profit by giving up the impossibility of ever going broke. (When trading according to Kelly or any fractional money management scheme we can never go broke. However, this is really a theoretical point because most traders will get fired as quickly for turning in a 90 percent drawdown as a 100 percent drawdown.) To see the general principle behind this trade-off we now look at Oscar's system, a progressive betting system first devised by a craps player in the 1950s (Wilson 1965).

A progressive betting system can never turn a negative edge into a positive edge. There is no magic. But the sizing algorithm can change certain aspects of the payout schedule. We have already seen this. We earlier mentioned the skewness generated by doubling bets after losses. Also, the fractional Kelly system traded lower returns for smaller drawdowns. Oscar wanted something entirely different. He just wanted every weekend he spent in Las Vegas to end with a small win. So he devised a progression that works exceedingly well in the short run.

There are two general types of betting progressions: positive progressions and negative progressions. With a positive progression, the general theory is that you raise your bets after wins, which means that your bigger bets are primarily funded by prior profits. The Kelly scheme is an example of a positive progression.

With a negative progression, you raise your bets after losses. This attempt to get back even more quickly is more dangerous, since a bad run of losses can wipe you out quickly. However, these schemes are seductive in that they allow you to win after a session in which you've lost more bets than you've won. Since your bets after losses are bigger bets, you don't have to win so many of them to come back. Or you lose so many consecutive bets that you go broke.

Many attempts have been made to combine the best features of these systems. Oscar's system was one. Oscar wanted to win just one unit. Each session started with a one-unit bet. If it won, he stopped. If he lost, the next bet was the same size. (So we take more risk than Kelly because our bets as a percentage of our bankroll grow as we lose.) After a win, the next bet would be one-unit higher than the last. No bet would ever be so large that it would take us over the target.

This system has been analyzed in detail, first by Wilson and then more extensively by Ethier (1996). The probability of success is shown in Table 8.2.

The problem is that in the rare instances that things go badly, they go horrendously badly. According to Wilson, in the 1960s Julian Braun ran a computer simulation of Oscar's system. He assumed the house had a betting limit of 500 units and that the probability of success for each bet was 244/495 (consistent with craps). In 280,000 trials there were 66 disasters where the gambler bumped up against the house limit. These situations lost an average of 13,000 units. So the laws of mathematics can't be cheated, but we can choose to push the disasters into the future rather than experience them continuously. To a certain degree, you can choose when to take your (inevitable) losses.

TABLE 8.2 The Probability of Reaching a Profit Target of one unit before having to quit when using Oscar's system, as a function of the win probability for a single trial, p, and the house limit M

M	$p = 9/19$	$p = 244/495$	$p = 1/2$
50	0.99078112	0.99620367	0.99734697
100	0.99464246	0.99841219	0.99904342
150	0.99587158	0.99902785	0.99947577
200	0.99646866	0.99930553	0.99965833
250	0.99681764	0.99946077	0.99975501
300	0.99704404	0.99955895	0.99981337
350	0.99720107	0.99962624	0.99985174

The situation for traders is different. Here we have a positive expectation. But we would still like to perform a similar trick to Oscar and somewhat smooth our stream of profits. We would also like to reduce our dependency on the long run. This was the problem addressed by Browne (1999, 2000). Specifically he found the dynamic strategy that maximized the probability of reaching a given wealth level in a specified time. He shows that the optimal fraction to invest with T, time left to reach the Goal B when current wealth is W, is given by

$$f^* = \frac{1}{\sigma\sqrt{T}} \frac{B\exp(-rT)}{W} n\left(N^{-1}\left(\frac{W\exp(rT)}{B}\right)\right) \tag{8.33}$$

where

$$n(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right) \tag{8.34a}$$

$$N(x) = \int_{=\infty}^{x} n(z)dz \tag{8.34b}$$

and r is the interest rate.

Browne shows that this sizing strategy is equivalent to the hedging strategy of a binary call. This argument is insightful and should help option traders get a feel for the behavior of the strategy. It may also help us think of ideas for generalizing the argument to more realistic scenarios, for example, where we have a stop placed on our bankroll as well as a target on the upside.

If we have a stock that evolves according to our normal model of GBM, the value of an option that pays B if we are above the strike, K, at time, T, is given by

$$C = B\exp(-r(T-t))N\left(\frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right) \tag{8.35}$$

(Haug 2007b).

The delta of this option is given by

$$\Delta = B\exp(-r(T-t))n\left(\frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right)\frac{1}{S\sigma\sqrt{T-t}} \tag{8.36}$$

Δ is the number of shares in the hedging portfolio, so at any time the value of the hedge is given by ΔxS. If we were actually short the option, our aim would be to hedge it so that we maximized the probability that we could pay off the claim at expiry. Or, at any given time our "wealth" is given by the expression in Equation 8.35 and we need to maximize the probability that our terminal wealth is given by B. The optimal policy in such a case is given by Equation 8.33, here with wealth $x = C(t, S)$. Making this substitution we obtain

$$f^* = \Delta S \tag{8.37}$$

So this strategy is indeed equivalent to the hedging strategy of the digital call where we consider the "stock price" to be our wealth, W. Figures 8.11 and 8.12 show the difference between this dynamic sizing strategy and the constant Kelly strategy. This assumes that we are trading a stock with a drift of 22 percent, an annualized volatility of 45 percent, and there is also a risk-free rate of 8 percent. In this case the Kelly ratio is 0.6914. Our goal is to make 50 percent in a session of 100 trading days (not completely unreasonable but very optimistic given the drift and volatility of the underlying). We can see that initially the Browne strategy is much more aggressive. We know that our goal requires us to take significant risk. Initially we trade at a leverage of 1.8, three times the leverage of the Kelly strategy. But as we get closer to our goal, we significantly dial down the risk. In practice, we would continually monitor our goals and make adjustments (this process is discussed in Chapter 9).

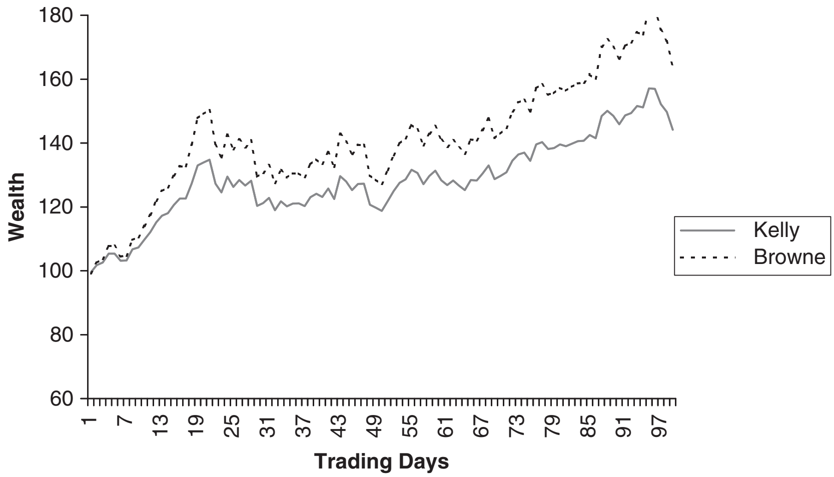

FIGURE 8.11 The Wealth Accumulated When Following the Browne and Kelly Strategies

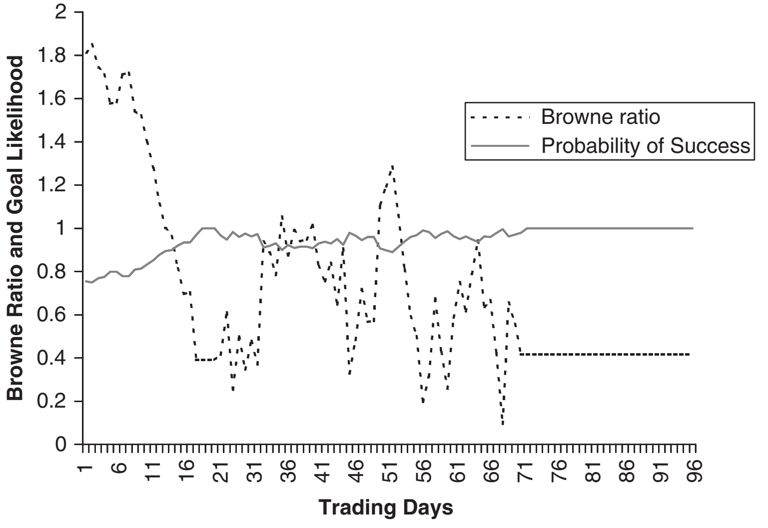

FIGURE 8.12 The Browne Ratio and the Probability of Reaching the Goal

TABLE 8.3 **Time Needed for Browne's Strategy to Beat Competing Strategies by 10 Percent**

Bankruptcy Probability	Time to Beat Cash by 10 Percent (years)	Time to Beat Stock by 10 Percent (years)
0.05	1.3	85
0.01	14	900
0.001	43	2,780
0.0001	73	4,774

He also shows that the expected time for this strategy to beat any other is given by

$$T = \left[\frac{N^{-1}(1-\alpha) - N^{-1}\left(\frac{1}{1-\varepsilon}\right)}{\sigma(f - f')} \right]^2 \qquad (8.38)$$

So now we can compare this strategy to Kelly's results given above.

These numbers are far better than for the Kelly strategy. Recall that for Kelly to beat the stock by 10 percent (at the 95 percent confidence level), it would take 10,286 years. Granted, trading using the Kelly criterion would leave us with no chance of bankruptcy but the significant improvement in the expected time to perform may well make this small risk worthwhile.

As we can see according to the criteria of expected time to dominate, this method is far better than Kelly. It is also riskier. Specifically, we expect to achieve our goal with probability V, and go bankrupt with probability $1 - V$, where V is given by

$$V = N\left(N^{-1}\left(\frac{W\exp(rT)}{B}\right) + \frac{\mu - r}{\sigma}\sqrt{T}\right) \tag{8.39}$$

and where

μZ is the expected drift of the asset.

V is also shown above in Figure 8.12.

As we said earlier, when utility functions are involved we need to decide what exactly we mean by *successful*.

The situations examined so far have been static. Either we have had an opportunity to place a single trade/bet and see how it fares or we have had the opportunity to invest in an asset with fixed drift and volatility. A more realistic situation is one where the parameters we are interested in are changing. In particular we have seen in Chapter 5 that implied volatility is a mean-reverting process. Considering these situations leads to a seemingly paradoxical situation.

Proebsting's paradox is the counterintuitive result that application of the Kelly criterion can seemingly lead to bankruptcy. As with most "paradoxes" this one can be resolved, but looking at the argument and its resolution is nevertheless worthwhile. The situation was first pointed out by Todd Proebsting in an e-mail to Edward Thorpe who wrote about it in *Wilmott Magazine* (Thorpe 2008).

Imagine a situation where a gambler is offered a bet with a 50 percent chance of winning. A win pays \$2 and a loss loses \$1. Equation 8.5 tells us that the Kelly fraction is 0.25. Now, before the game is played, the bettor is offered another bet on the same event. This one pays \$5 if it wins and still only can lose \$1. We need to determine the fraction of bankroll, f, to allocate to this new opportunity. After the bet is resolved the gambler's bankroll will either be $1.5\,W_0 + 5f\,W_0$, if he wins, or $0.75\,W_0 - f\,W_0$ if he loses. So we need to maximize

$$0.5\,\ln(1.5W_0 + 5fW_0) + 0.5\ln(0.75W_0 + fW_0) \tag{8.40}$$

This gives $f = 0.225$.

So the total fraction (the total bet on first the 2/1 bet and then the 5/1 bet) is 0.425. The paradox is that if only the 5/1 bet was offered the Kelly fraction would be 0.4. So application of the Kelly criterion leads us to be *more* when some of the bet is at *worse* odds.

Even more alarmingly, if the gambler is offered better and better odds he will eventually bet all of his money (in the limit). This gives him a

100 percent chance of going bankrupt and was precisely the situation that the Kelly criterion was designed to avoid.

Proebsting's paradox is important because this situation occurs all the time in trading. Prices change continuously and our edge (odds) fluctuates accordingly. The most satisfactory resolution of the paradox was given by Aaron Brown (also in an e-mail to Ed Thorpe). In a rare case of finance theory illuminating a gambling issue rather than vice versa, Brown made use of the idea of marking to market.

Specifically when the gambler is offered the 5/1 bet it means that on a marked to market basis his 2/1 bet is a loser. That bet may not have settled but it is still a loser as we can now get a better payoff for the same cost. And, as any derivatives trader should know, it is the daily P/L that matters, not what would happen if the trade was held until expiration. What is the change in the gambler's bankroll when the odds move from 2/1 to 5/1? Alternatively, how much does he have to pay to get at the current market price of 5/1, a bet that pays 2 for a win? This is the value of γ that satisfies

$$(2-\gamma) = 5(1+\gamma)$$
$$\gamma = -0.5$$

that is, entering a 2/1 bet for 0.25 of our bankroll and then having the odds change to 5/1 is the same as not having bet, having the bankroll change by $-0.5 \times 0.25W = -0.125W$, and then betting at 5/1 odds.

Generally if we bet f^X on a bet with a payout of $X/1$, then we get offered $Y/1$ odds (where $Y > B$) the mark-to-market wealth (which we should use to calculate the new betting fraction) is

$$W = \frac{1+\bar{x}}{1+Y} W_0 \tag{8.41}$$

where $\bar{x}$ is the weighted average of the odds

$$\bar{x} = f^x X \quad + \quad (1-f^x)Y \tag{8.42}$$

We will see that trading a mean-reverting process is similar. As the deviations from fair value get larger we will do more of the trade. This leads to the rule of thumb: When in trouble, double. But this is true only to a point. In a continuous trading setting, as we get better and better prices we will be losing money on the position we have already established. Eventually these losses will have depleted our bankroll to the point where we actually want to have a smaller position on. This is consistent with how market makers are told to size their trades when they have things pushed against them. The old rule would have a market maker sell 100 at his first level, then back up and sell 200 at the next level, back up again and sell 300 at the next level. But if things get pushed further from "fair value" he would start to buy the position back. This isn't necessarily because his estimate of what is fair might have changed (although by this point it probably will have done) but because he has lost money and the short 600-lot position is too large a function of his new account size.

We will set up a simple model that captures the basic features of what we are trading, and try to generalize the Kelly argument. The simplest mean-reverting model is the single parameter Ornstein-Uhlenbeck process, governed completely by its speed of reversion, μ. Also let's assume for simplicity that we have normalized the underlying, S, so it has a mean of zero and a standard deviation of one.

$$dS = -\mu S dt + \sqrt{2\mu} dZ \tag{8.43}$$

As with GBM these asset paths look very noisy and just looking at them would be a very bad way to estimate the true reversion speed. Actually, visual inspection wouldn't even be enough to see that these were mean reverting. For example, the three paths shown in Figure 8.13 all simulate five years of daily prices with a reversion speed of 100 percent.

The optimal asset allocation for such a process has been studied (Boguslavsky and Boguslavskaya 2004; Liu and Longstaff 2004). They showed that if we try to generalize the Kelly approach and maximize the expected value of the logarithm of wealth the solution is that we should hold

$$-W \times \sigma/2 \tag{8.44}$$

in the asset.

So if we are risking \$100 and the price of the asset is 1.8 standard deviations from its mean (i.e., $\sigma = 1.8$) then we will have a short position of $100 \times 1.8/2 = 90$. In practical terms this means we should have on a position such that if the spread goes back to the mean we should make \$90.

Figures 8.14 and 8.15 show the results of a "trading session" where we size our trades according to this rule.

There are a number of important things to note about this deceptively simple result.

- Wealth is equally important as the attractiveness of the trade (σ and W are interchangeable in the result).
- The maximum position is at $\sigma = 2^{1/2}$. Beyond this the effect of losing wealth dominates. The trades are now better but our bankroll is now also smaller.

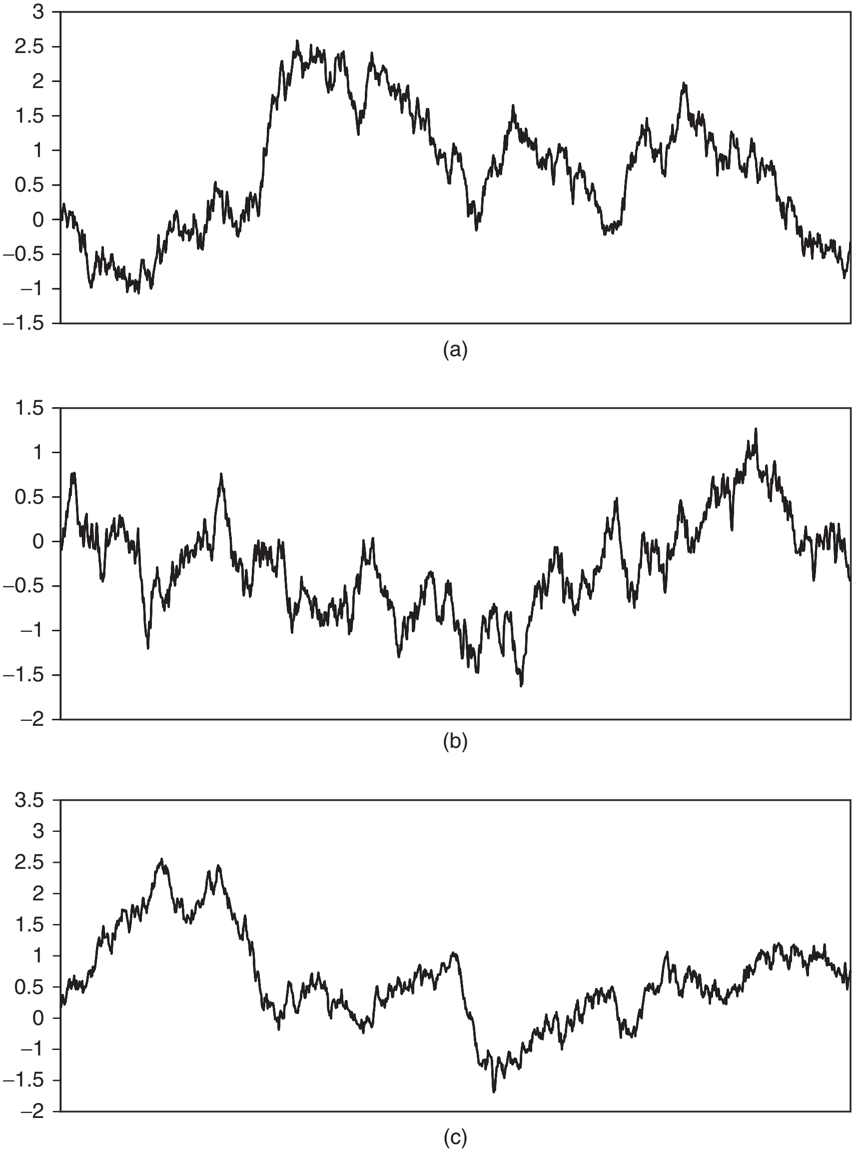

FIGURE 8.13 Three Different Instances of an Asset Path Generated by the Same Mean-Reverting Process

- At deviations smaller than $2^{1/2}$, we add to trades as they go against us. Here the extra edge in the trade dominates the fact that we are losing money on our position.
- This result is independent of the reversion speed. High reversion rates are good, but only because we get to do more trades in the same period of time.

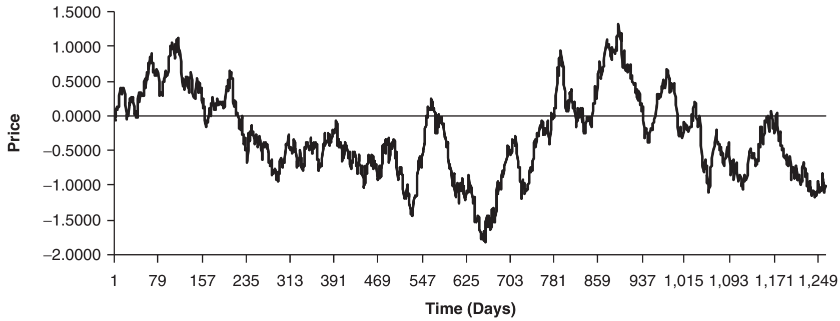

FIGURE 8.14 The Mean Reverting Asset Price

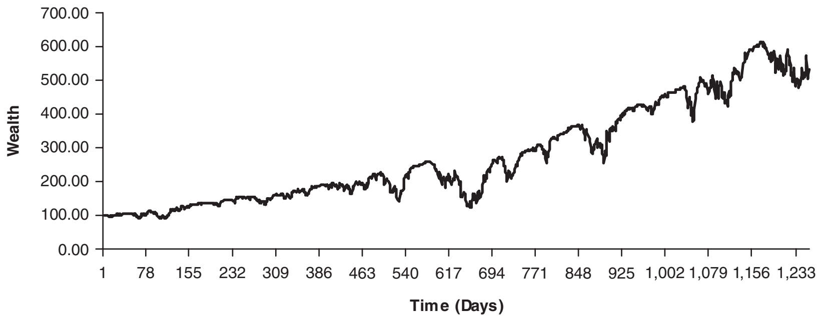

FIGURE 8.15 The Wealth Generated by Trading According to Equation 8.44

As with all strategies that maximize the expectation of the logarithm of wealth, this is an aggressive money management scheme. To some extent this can be partially managed by careful choice of W_0 (where we will apply partial instead of full Kelly). But a further danger exists here. We have assumed that the process is governed by an Ornstein-Uhlenbeck process with normal innovations. The real processes we deal with in finance will generally have fatter tails than implied by the normal distribution. Things will not work so well in this case. Figures 8.16 and 8.17 show the results of a trading session where the distribution is logistic rather than normal. The important difference between this example and that of Figures 8.12 and 8.13 is that here we have an excess kurtosis of 9. The results are clearly affected for the worse.

There doesn't seem to be any published work on optimal positioning for these types of processes, but some intuition can be gained by running some simulations. Fat-tailed distributions also have thin middles. So in addition to the larger proportion of large moves we can also expect more small moves. This would seem to suggest that modifying the strategy by trading more aggressively when entering looking for small moves then exiting

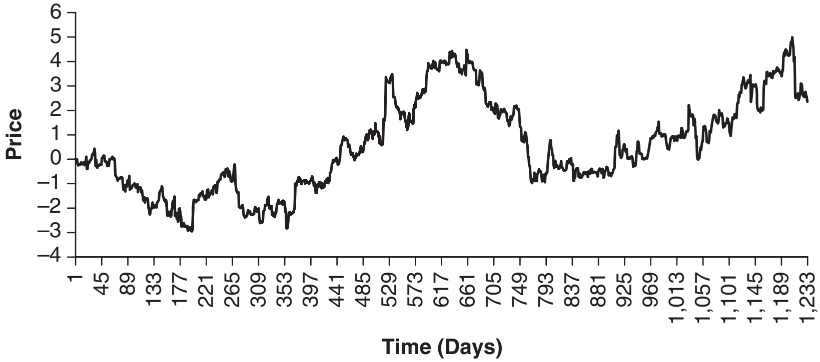

FIGURE 8.16 The Mean-Reverting Asset Price

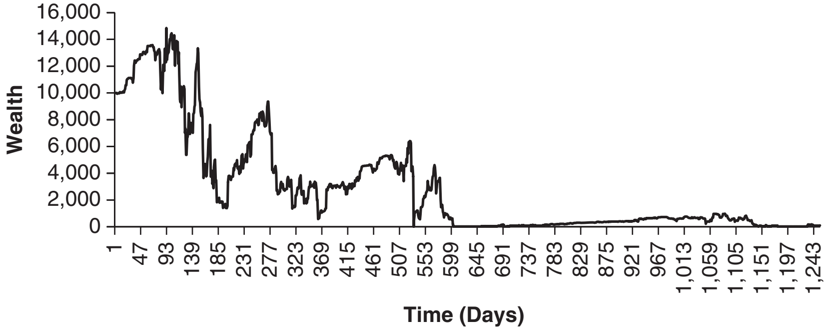

FIGURE 8.17 The Wealth Generated by Trading According to Equation 8.44

more quickly when things move against us. Numerical experiments seem to support this.

It is important to emphasize the dangerous game we are playing here. This theory shows that when trading a mean-reverting process we should (at first) add to our position as it goes against us. This clearly can be dangerous. The danger has nothing to do with such unsupported rules as "only losers add to losers." In fact, the theory shows that the optimal trading rule in this instance is to add to losers. Further, this rule can be enunciated, tested, and modified in a way that vague general assertions cannot. The real danger is that the system may fundamentally change so that the "mean" we will eventually revert to is totally different to the one we started with. This is why we need to fundamentally evaluate why the initial trade has gone against us. Imagine we are short a stock at 20 percent volatility because we think the fair value is 13 percent. If it gets choppy because a few large orders enter the market on an otherwise slow day and implied volatility rises to 22 percent, we are probably justified in selling more. However, consider the example of InterOil Corporation (IOC). At 2 P.M. EST on June 26, 2007, huge sell orders hit the market and drove the stock price down from \$40.2 to \$26.5. There was no news about the company on any of the major news wires. July implied volatility jumped from about 94 percent to 120 percent. In this case a fundamental revaluation of the company had occurred. We have no news or analysis to consider and huge volume was being transacted (6 million shares in an hour in a stock with an average daily volume of 600,000). In this case to sell more volatility would be foolhardy and irresponsible. So each case needs to be considered in context.

This is obviously an area where experience of a particular market can be an asset. But there is a good way to use experience and a bad way. The good way is to use your knowledge defensively. You should actively look for things that are out of place with what you have seen in the past and then be extra cautious. The bad way is to overfit to past data. If you have never seen deviations as large before, that doesn't necessarily mean that this is the best trade you have ever done. This could well mean your past experience is now irrelevant.

Selling a rising market and buying a falling market is the replication strategy for a short option. In Chapter 4, we saw that forecast volatility is generally below implied volatility. Part of the reason for this was that in selling implied volatility we were selling insurance against events that have never occurred before. That is also the case when scaling into any mean-reverting asset. You have to be aware of this and be prepared to stop scaling in, even when it can look better than it ever has before.

In practice, it will generally be difficult, or even impossible, to maintain the perfect level of positioning. Illiquidity, transaction costs, noncontinuous trading, position limits, and order entry restrictions mean that at best we will need to use Equation 8.44 as a guide. It might also be helpful to have a discrete rule to guide us in putting on the first portion of the position. Let's try a simple argument for maximizing our total profit given, that we have to choose one point for entering the trade. This and other rules for selecting entry points in mean reverting processes were covered in Vidyamurthy (2004).

In this toy model we assume that the deviations of the asset price from its mean value are normally distributed and independent. So at any time we just draw a number from a normal distribution to find the deviation and this is independent of the previous values. For a normally distributed process, the probability at any time that we have deviated by more than S from the mean is just the integral of the process. This is equal to $1 - N(S)$ (where N [.] is the cumulative normal distribution function). So in, T, time steps we can expect to have $T(1 - N(S))$ times, where the asset price has a deviation greater than or equal to S. The normal distribution is symmetric so we have an equal number of times where the spread is less than or equal to $-S$. So in T time steps, we will have traded $2T(1 - N(S))$ times. Each of these gives a profit of S. So the total profit is given by

$$2T\,S(1 - N(S)) \tag{8.45}$$

To find the maximum of this function with respect to S, we differentiate and set the result equal to zero. This gives the result

$$S_{\max} = 0.75\,\sigma \tag{8.46}$$

Figure 8.18 shows the shape of the theoretical distribution of the P/L as a function of the entry level.

How different will trading a real process be? A true financial process will differ from this ideal in three important ways.

1. The distribution will have fat tails.
2. There will be different behavior associated with declines and increases (for example, the Volatility Index (VIX) is mean-reverting but is prone to have larger moves up than down).
3. The standard deviation of the process will not be constant.

But rather than use a more complex model, we now look at trading a mean-reverting product with various entry levels and see how the P/L varies in practice. We simulate trading the VIX. We know this is a mean-reverting process. Here we follow a simple Bollinger band rule. We buy or sell the VIX after it has deviated by a certain amount from its moving average. As can be seen, the deviation from the simple moving average is somewhat

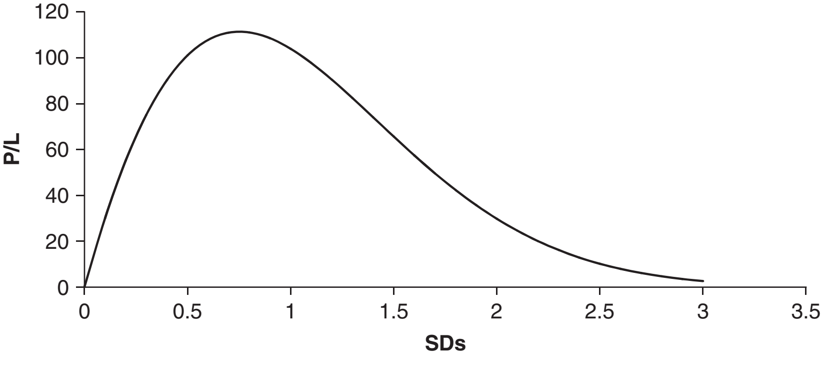

FIGURE 8.18 The Shape of the P/L Distribution as a Function of Entry Level

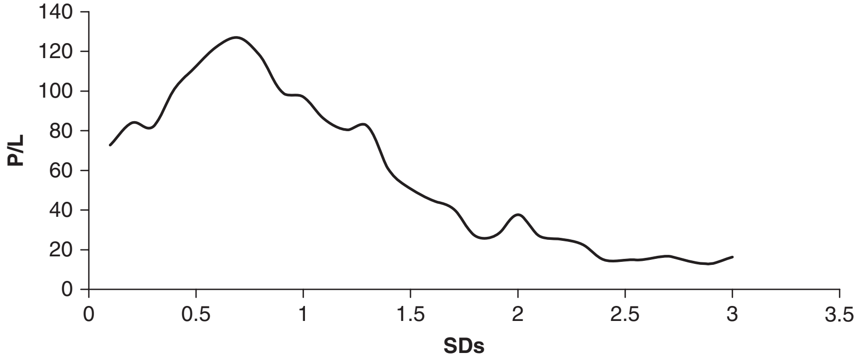

FIGURE 8.19 The Shape of the P/L Distribution for the VIX Trade

normal looking but clearly is fat tailed and skewed. We could address the skewness issue by having different bands for buying and selling but that is not the point of this exercise. This is not meant to be a realistic trading idea. It is just intended to show that Equation 8.46 has some applicability to a situation that we know does not follow the necessary simplifying assumptions.

We can see in Figure 8.19 that the peak of the P/L function is close to the theoretical point of 0.75 standard deviations. So aggressively scalping can produce more profits. But note that the left-hand end of the curve tails off much faster so it is probably safer to err on the side of caution and trade slightly less often than optimal.

Summary

Although trade sizing can't turn a losing trade into a winning one, our choices still have a significant impact on our profitability, variance, and drawdowns. It is important that our sizing scheme and parameters are established before we start trading, and then only adjusted after very careful consideration. It is easy to overreact to atypically good or bad runs of results.

- Before you can choose a money management scheme you must be clear what you are trying to accomplish. Your monetary goals, time constraints, and maximum tolerable drawdowns need to be fully specified in advance.
- When there is more edge, trade bigger.
- When there is more variance or uncertainty, trade smaller.

- The Kelly scheme will eventually overwhelm all others.
- The Browne scheme is useful for hitting specific targets.
- When trading volatility we have to be prepared to do more as a trade initially goes against us.
- Adding arbitrary price based stops to a trading system is a poor idea. We should exit our trades when we are wrong. Having volatility move against us may not indicate that we are wrong at all.
- A good rule of thumb is that a trade should be big enough that the profits mean something, but not so big that the losses are catastrophic. If this optimum size can't be found, the trade probably doesn't have enough edge to begin with.