CHAPTER 5

Characteristics of Optimal *f*

OPTIMAL *f* FOR SMALL TRADERS JUST STARTING OUT

How does a very small account, an account that is going to start out trading one contract, use the optimal f approach? One suggestion is that such an account start out by trading one contract not for every optimal f amount in dollars (biggest loss/$-f$), but rather that the drawdown and margin must be considered in the initial phase. The amount of funds allocated toward the first contract should be the greater of the optimal f amount in dollars or the margin plus the maximum historic drawdown (on a one-unit basis):

$$A = \text{MAX}\{(\text{Biggest Loss}/-f), (\text{Margin} + \text{ABS}(\text{Drawdown}))\} \qquad (5.01)$$

where:
A = The dollar amount to allocate to the first contract.
f = The optimal f (0 to 1).
Margin = The initial speculative margin for the given contract.
Drawdown = The historic maximum drawdown.
MAX{ } = The maximum value of the bracketed values.
ABS( ) = The absolute value function.

With this procedure an account can experience the maximum drawdown again and still have enough funds to cover the initial margin on another trade. Although we cannot expect the worst-case drawdown in the future not to exceed the worst-case drawdown historically, it is rather unlikely that we will start trading right at the beginning of a new historic drawdown.

A trader utilizing this idea will then subtract the amount in Equation (5.01) from his or her equity each day. With the remainder, he or she will then divide by (Biggest Loss/$-f$). The answer obtained will be rounded down to the integer, and 1 will be added. The result is how many contracts to trade.

An example may help clarify. Suppose we have a system where the optimal f is .4, the biggest historical loss is $-$\$3,000, the maximum drawdown was $-$\$6,000, and the margin is \$2,500. Employing Equation (5.01) then:

$$\begin{aligned} A &= \text{MAX}\{(-\$3,000/-.4), (\$2,500 + \text{ABS}(-\$6,000))\} \\ &= \text{MAX}\{(\$7,500), (\$2,500 + \$6,000)\} \\ &= \text{MAX}\{\$7,500, \$8,500)\} \\ &= \$8,500 \end{aligned}$$

We would thus allocate \$8,500 for the first contract. Now suppose we are dealing with \$22,500 in account equity. We therefore subtract this first contract allocation from the equity:

$$\$22,500 - \$8,500 = \$14,000$$

We then divide this amount by the optimal f in dollars:

$$\$14,000/\$7,500 = 1.867$$

Then we take this result down to the integer:

$$\text{INT}(1.867) = 1$$

and add 1 to the result (the one contract represented by the \$8,500 we have subtracted from our equity):

$$1 + 1 = 2$$

We therefore would trade two contracts. If we were just trading at the optimal f level of one contract for every \$7,500 in account equity, we would have traded three contracts (\$22,500/\$7,500). As you can see, this technique can be utilized no matter how large an account's equity is (yet the larger the equity, the closer the two answers will be). Further, the larger the equity, the less likely it is that we will eventually experience a drawdown that will have us eventually trading only one contract. For smaller accounts, or for accounts just starting out, this is a good idea to employ.

THRESHOLD TO GEOMETRIC

Here is another good idea for accounts just starting out, one that may not be possible if you are employing the technique just mentioned. This technique makes use of another by-product calculation of optimal f called the *threshold to geometric*. The by-products of the optimal f calculation include calculations, such as the TWR, the geometric mean, and so on, that were derived in obtaining the optimal f, and that tell us something about the system. The threshold to the geometric is another of these by-product calculations. Essentially, *the threshold to geometric tells us at what point we should switch over to fixed fractional trading, assuming we are starting out constant-contract trading.*

Refer back to the example of a coin toss where we win \$2 if the toss comes up heads and we lose \$1 if the toss comes up tails. We know that our optimal f is .25, or to make one bet for every \$4 we have in account equity. If we are starting out trading on a constant-contract basis, we know we will average \$.50 per unit per play. However, if we start trading on a fixed fractional basis, we can expect to make the geometric average trade of \$.2428 per unit per play.

Assume we start out with an initial stake of \$4, and therefore we are making one bet per play. Eventually, when we get to \$8, the optimal f would have us step up to making two bets per play. However, two bets times the geometric average trade of \$.2428 is \$.4856. Wouldn't we be better off sticking with one bet at the equity level of \$8, whereby our expectation per play would still be \$.50? The answer is "Yes." The reason is that the optimal f is figured on the basis of contracts that are infinitely divisible, which may not be the case in real life.

We can find that point where we should move up to trading two contracts by the formula for the threshold to the geometric, T:

$$\text{T} = \text{AAT/GAT} * \text{Biggest Loss}/{-f} \tag{5.02}$$

where:
T = The threshold to the geometric.
AAT = The arithmetic average trade.
GAT = The geometric average trade.
f = The optimal f (0 to 1).

In our example of the 2-to-1 coin toss:

$$\begin{aligned} \text{T} &= .50/.2428 * -1/-.25 \\ &= 8.24 \end{aligned}$$

Therefore, we are better off switching up to trading two contracts when our equity gets to \$8.24 rather than \$8. Figure 5.1 shows the threshold to the

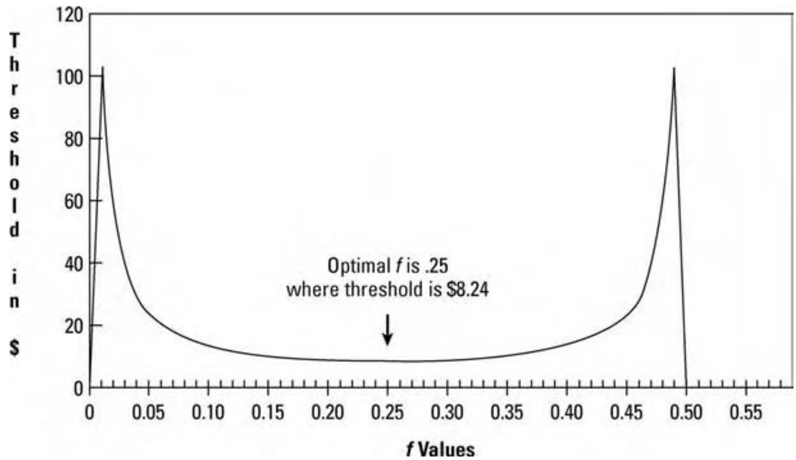

FIGURE 5.1 Threshold to the geometric for 2:1 coin toss

geometric for a game with a 50% chance of winning \$2 and a 50% chance of losing \$1.

Notice that the trough of the threshold to the geometric curve occurs at the optimal *f*. This means that since the threshold to the geometric is the optimal level of equity to go to trading two units, you go to two units at the lowest level of equity, optimally, when incorporating the threshold to the geometric at the optimal *f*.

Now the question is, "Can we use a similar approach to know when to go from two cars to three cars?" Also, "Why can't the unit size be 100 cars starting out, assuming you are starting out with a large account, rather than simply a small account starting out with one car?" To answer the second question first, it is valid to use this technique when starting out with a unit size greater than one. However, it is valid only if you do *not* trim back units on the downside before switching into the geometric mode. The reason is that before you switch into the geometric mode you are assumed to be trading in a constant-unit size.

Assume you start out with a stake of 400 units in our 2-to-1 coin-toss game. Your optimal *f* in dollars is to trade one contract (make one bet) for every \$4 in equity. Therefore, you will start out trading 100 contracts (making 100 bets) on the first trade. Your threshold to the geometric is at \$8.24, and therefore you would start trading 101 contracts at an equity level of \$404.24. You can convert your threshold to the geometric, which is computed on the basis of advancing from one contract to two, as:

$$\text{Converted T} = \text{EQ} + \text{T} - (\text{Biggest Loss}/-f) \tag{5.02a}$$

where: EQ = The starting account equity level.
T = The threshold to the geometric for going from one car to two.
f = The optimal f (0 to 1).

Therefore, since your starting account equity is \$400, your T is \$8.24, your biggest loss −\$1, and your f is .25:

$$\begin{aligned}\text{Converted T} &= 400 + 8.24 - (-1/-.25)\\ &= 400 + 8.24 - 4\\ &= 404.24\end{aligned}$$

Thus, you would progress to trading 101 contracts (making 101 bets) if and when your account equity reached \$404.24. We will assume you are trading in a constant-contract mode until your account equity reaches \$404.24, at which point you will begin the geometric mode. Therefore, until your account equity reaches \$404.24, you will trade 100 contracts on the next trade regardless of the remaining equity in your account. If, after you cross the geometric threshold (that is, after your account equity hits \$404.24), you suffer a loss and your equity drops below \$404.24, you will go back to trading on a constant 100-contract basis if and when you cross the geometric threshold again.

This inability to trim back contracts on the downside when you are below the geometric threshold is the drawback to using this procedure when you are at an equity level of trading more than two contacts. If you are only trading one contract, the geometric threshold is a very valid technique for determining at what equity level to start trading two contracts (since you cannot trim back any further than one contract should you experience an equity decline). However, it is not a valid technique for advancing from two contracts to three, because the technique is predicated upon the fact that you are currently trading on a constant-contract basis. That is, if you are trading two contracts, unless you are willing not to trim back to one contract if you suffer an equity decline, the technique is not valid, and likewise if you start out trading 100 contracts. You could do just that (not trim back the number of contracts you are presently trading if you experience an equity decline), in which case the threshold to the geometric, or its converted version in Equation (5.02a), would be the valid equity point to add the next contract. The problem with doing this (not trimming back on the downside) is that you will make less (your TWR will be less) in an asymptotic sense. You will not make as much as if you simply traded the full optimal f. Further, your drawdowns will be greater and your risk of ruin higher. Therefore, the threshold to the geometric is beneficial only if you are starting out in the lowest denomination of bet size (one contract) and advancing to two, and it is a benefit only if the arithmetic average trade is more than twice the size of the geometric average trade. Furthermore, it is beneficial to use only when you cannot trade fractional units. In Chapter 10 we will see that the concept of a "Threshold" to the geometric is a precursor to the larger notion of Continuous Dominance.

ONE COMBINED BANKROLL VERSUS SEPARATE BANKROLLS

Some very important points regarding fixed fractional trading must be covered before we discuss the parametric techniques. First, when trading more than one market system simultaneously, you will generally do better in an asymptotic sense using only one combined bankroll from which to figure your contract sizes, rather than separate bankrolls for each.

It is for this reason that we "recapitalize" the subaccounts on a daily basis as the equity in an account fluctuates. What follows is a run of two similar systems, System A and System B. Both have a 50% chance of winning, and both have a payoff ratio of 2:1. Therefore, the optimal *f* dictates that we bet \$1 for every \$4 units in equity. The first run we see shows these two systems with positive correlation to each other. We start out with \$100, splitting it into two subaccount units of \$50 each. After a trade is registered, it affects only the cumulative column for that system, as each system has its own separate bankroll. The size of each system's separate bankroll is used to determine bet size on the subsequent play:

System A			System B		
Trade	**P&L**	**Cumulative**	**Trade**	**P&L**	**Cumulative**
		50.00			50.00
2	25.00	75.00	2	25.00	75.00
−1	−18.75	56.25	−1	−18.75	56.25
2	28.13	84.38	2	28.13	84.38
−1	−21.09	63.28	−1	−21.09	63.28
2	31.64	94.92	2	31.64	94.92
−1	−23.73	71.19	−1	−23.73	71.19
		−50.00			−50.00
Net Profit		21.19140			21.19140
	Total net profit of the two banks =				\$42.38

Now we will see the same thing, only this time we will operate from a combined bank starting at 100 units. Rather than betting $1 for every $4 in the combined stake for each system, we will bet $1 for every $8 in the combined bank. Each trade for either system affects the combined bank, and it is the combined bank that is used to determine bet size on the subsequent play:

System A		System B		
Trade	**P&L**	**Trade**	**P&L**	**Combined Bank**
				100.00
2	25.00	2	25.00	150.00
−1	−18.75	−1	−18.75	112.50
2	28.13	2	28.13	168.75
−1	−21.09	−1	−21.09	126.56
2	31.64	2	31.64	189.84
−1	−23.73	−1	−23.73	142.38
				−100.00
	Total net profit of the combined bank =			$42.38

Notice that using either a combined bank or a separate bank in the preceding example shows a profit on the $100 of $42.38. Yet what was shown is the case where there is positive correlation between the two systems. Now we will look at negative correlation between the same two systems, first with both systems operating from their own separate bankrolls:

System A			System B		
Trade	**P&L**	**Cumulative**	**Trade**	**P&L**	**Cumulative**
		50.00			50.00
2	25.00	75.00	−1	−12.50	37.50
−1	−18.75	56.25	2	18.75	56.25
2	28.13	84.38	−1	−14.06	42.19
−1	−21.09	63.28	2	21.09	63.28
2	31.64	94.92	−1	−15.82	47.46
−1	−23.73	71.19	2	23.73	71.19
		−50.00			−50.00
Net Profit		21.19140			21.19140
	Total net profit of the two banks =				$42.38

As you can see, when operating from separate bankrolls, both systems net out making the same amount regardless of correlation. However, with the combined bank:

System A		System B		
Trade	**P&L**	**Trade**	**P&L**	**Combined Bank**
				100.00
2	25.00	−1	−12.50	112.50
−1	−14.06	2	28.12	126.56
2	31.64	−1	−15.82	142.38
−1	−17.80	2	35.59	160.18
2	40.05	−1	−20.02	180.20
−1	−22.53	2	45.00	202.73
				−100.00
	Total net profit of the combined bank =			$102.73

With the combined bank, the results are dramatically improved. *When using fixed fractional trading you are best off operating from a single combined bank.*

TREAT EACH PLAY AS IF INFINITELY REPEATED

The next axiom of fixed fractional trading regards maximizing the current event as though it were to be performed an infinite number of times in the future. We have determined that for an independent trials process, you should always bet that f which is optimal (and constant) and likewise when there is dependency involved, only with dependency f is not constant.

Suppose we have a system where there is dependency in like begetting like, and suppose that this is one of those rare gems where the confidence limit is at an acceptable level for us, that we feel we can safely assume that there really is dependency here. For the sake of simplicity we will use a payoff ratio of 2:1. Our system has shown that, historically, if the last play was a win, then the next play has a 55% chance of being a win. If the last play was a loss, our system has a 45% chance of the next play's being a loss. Thus, if the last play was a win, then from the Kelly formula, Equation (4.03), for finding the optimal *f* (since the payoff ratio is Bernoulli distributed):

$$\begin{aligned} f &= ((2+1) * .55 - 1)/2 \\ &= (3 * .55 - 1)/2 \\ &= .65/2 \\ &= .325 \end{aligned}$$

After a losing play, our optimal *f* is:

$$\begin{aligned} f &= ((2+1) * .45 - 1)/2 \\ &= (3 * .45 - 1)/2 \\ &= .35/2 \\ &= .175 \end{aligned}$$

Now dividing our biggest losses (−1) by these negative optimal *f*s dictates that we make one bet for every 3.076923077 units in our stake after a win, and make one bet for every 5.714285714 units in our stake after a loss. In so doing we will maximize the growth over the long run. Notice that we treat each individual play as though it were to be performed an infinite number of times.

Notice in this example that betting after both the wins and the losses still has a positive mathematical expectation individually. What if, after a loss, the probability of a win was .3? In such a case, the mathematical expectation is negative, hence there is no optimal *f* and as a result you shouldn't take this play:

$$\begin{aligned} \text{ME} &= (.3 * 2) + (.7 * -1) \\ &= .6 - .7 \\ &= -.1 \end{aligned}$$

In such circumstances, you would bet the optimal amount only after a win, and you would not bet after a loss. If there is dependency present, you must segregate the trades of the market system based upon the dependency and treat the segregated trades as separate market systems.

The same principle, namely that *asymptotic growth is maximized if each play is considered to be performed an infinite number of times into the future*, also applies to simultaneous wagering (or trading a portfolio). Consider two betting systems, A and B. Both have a 2:1 payoff ratio, and both win 50% of the time. We will assume that the correlation coefficient between the two systems is zero, but that is not relevant to the point being illuminated here. The optimal *f*s for both systems (if they were being traded alone, rather than simultaneously) are .25, or to make one bet for every four units in equity. The optimal *f*s for trading both systems simultaneously are .23, or one bet for every 4.347826087 units in account equity.[1] System B trades only two-thirds of the time, so some trades will be done when the two systems are not trading simultaneously. This first sequence is demonstrated with a starting combined bank of 1,000 units, and each bet for each system is performed with an optimal *f* of one bet per every 4.347826087 units:

	A		B	Combined Bank
				1,000.00
−1	−230.00			770.00
2	354.20	−1	−177.10	947.10
−1	−217.83	2	435.67	1,164.93
2	535.87			1,700.80
−1	−391.18	−1	−391.18	918.43
2	422.48	2	422.48	1,763.39

Next, we see the same exact thing, the only difference being that when A is betting alone (i.e., when B does not have a bet at the same time as A), we make one bet for every four units in the combined bank for System A, since that is the optimal *f* on the single, individual play. On the plays where the bets are simultaneous, we are still betting one unit for every 4.347826087 units in account equity for both A and B. Notice that in so doing we are taking each bet, whether it is individual or simultaneous, and applying that optimal *f* which would maximize the play as though it were to be performed an infinite number of times in the future.

	A		B	Combined Bank
				1,000.00
−1	−250.00			750.00
2	345.00	−1	−172.50	922.50
−1	−212.17	2	424.35	1,134.67
2	567.34			1,702.01
−1	−391.46	−1	−391.46	919.09
2	422.78	2	422.78	1,764.65

As can be seen, there is a slight gain to be obtained by doing this and the more trades that elapse, the greater the gain. Although we are not yet

[1]The method we are using here to arrive at these optimal bet sizes is described later in the text in Chapter 9.

discussing multiple simultaneous plays (i.e., "portfolios"), we invoke them here to illuminate the point. The same principle applies to trading a portfolio where not all components of the portfolio are in the market all the time. You should trade at the optimal levels for the combination of components (or single component) that results in the optimal growth as though that combination of components (or single component) were to be traded an infinite number of times in the future.

EFFICIENCY LOSS IN SIMULTANEOUS WAGERING OR PORTFOLIO TRADING

Let's again return to our 2:1 coin-toss game. Let's again assume that we are going to play two of these games, which we'll call System A and System B, simultaneously and that there is zero correlation between the outcomes of the two games. We can determine our optimal *f*s for such a case as betting one unit for every 4.347826 in account equity when the games are played simultaneously. When starting with a bank of 100 units, notice that we finish with a bank of 156.86 units:

	System A		System B		
	Trade	**P&L**	**Trade**	**P&L**	**Bank**
Optimal *f* is 1 unit for every 4.347826 in equity:					
					100.00
	−1	−23.00	−1	−23.00	54.00
	2	24.84	−1	−12.42	66.42
	−1	−15.28	2	30.55	81.70
	2	37.58	2	37.58	156.86

Now let's consider System C. This would be the same as Systems A and B, only we're going to play this game alone, without another game going simultaneously. We're also going to play it for eight plays—as opposed to the previous endeavor, where we played two games for four simultaneous plays. Now our optimal *f* is to bet one unit for every four units in equity. What we have is the same eight outcomes as before, but a different, better end result:

	System C		
	Trade	P&L	Bank
Optimal *f* is 1 unit for every 4.00 in equity:			
			100.00
	−1	−25.00	75.00
	2	37.50	112.50
	−1	−28.13	84.38
	2	42.19	126.56
	2	63.28	189.84
	2	94.92	284.77
	−1	−71.19	213.57
	−1	−53.39	160.18

The end result here is better not because the optimal *f*s differ slightly (both are at their respective optimal levels), but because there is a small efficiency loss involved with simultaneous wagering. *This inefficiency is the result of not being able to recapitalize your account after every single wager as you could betting only one market system.* In the simultaneous two-bet case, you can recapitalize only three times, whereas in the single eight-bet case you recapitalize seven times. Hence, the efficiency loss in simultaneous wagering (or in trading a portfolio of market systems).

We just witnessed the case where the simultaneous bets were not correlated. Let's look at what happens when we deal with positive (+1.00) correlation:

	System A		System B		
	Trade	P&L	Trade	P&L	Bank
Optimal *f* is 1 unit for every 8.00 in equity:					
					100.00
	−1	−12.50	−1	−12.50	75.00
	2	18.75	2	18.75	112.50
	−1	−14.06	−1	−14.06	84.38
	2	21.09	2	21.09	126.56

Notice that after four simultaneous plays where the correlation between the market systems employed is +1.00, the result is a gain of 126.56 on a starting stake of 100 units. This equates to a TWR of 1.2656, or a geometric mean, a growth factor per play (even though these are combined plays) of 1.2656 ^ (1/4) = 1.06066.

Now refer back to the single-bet case. Notice here that after four plays, the outcome is 126.56, again on a starting stake of 100 units. Thus, the geometric mean of 1.06066. This demonstrates that the rate of growth is the same when trading at the optimal fractions for perfectly correlated markets. As soon as the correlation coefficient comes down below +1.00, the rate of growth increases. Thus, we can state that *when combining market systems, your rate of growth will never be any less than with the single-bet case, no matter how high the correlations are, provided that the market system being added has a positive arithmetic mathematical expectation.*

Recall the first example in this section, where there were two market systems that had a zero correlation coefficient between them. This market system made 156.86 on 100 units after four plays, for a geometric mean of (156.86/100) ^ (1/4) = 1.119. Let's now look at a case where the correlation coefficients are −1.00. Since there is never a losing play under the following scenario, the optimal amount to bet is an infinitely high amount (in other words, bet one unit for every infinitely small amount of account equity). But, rather than getting that greedy, we'll just make one bet for every four units in our stake so that we can make the illustration here:

	System A		System B		
	Trade	**P&L**	**Trade**	**P&L**	**Bank**
Optimal *f* is 1 unit for every 0.00 in equity (shown is 1 for every 4):					
					100.00
	−1	−12.50	2	25.00	112.50
	2	28.13	−1	−14.06	126.56
	−1	−15.82	2	31.64	142.38
	2	35.60	−1	−17.80	160.18

There are two main points to glean from this section. The first is that there is a small efficiency loss with simultaneous betting or portfolio trading, a loss caused by the inability to recapitalize after every individual play. The second point is that combining market systems, provided they have a positive mathematical expectation, and even if they have perfect positive correlation, never decreases your total growth per time period. However, as you continue to add more and more market systems, the efficiency loss becomes considerably greater. If you have, say, 10 market systems and they all suffer a loss simultaneously, that loss could be terminal to the account, since you have not been able to trim back size for each loss as you would have had the trades occurred sequentially.

Therefore, we can say that there is a gain from adding each new market system to the portfolio provided that the market system has a correlation coefficient less than one and a positive mathematical expectation, or a negative expectation but a low enough correlation to the other components in the portfolio to more than compensate for the negative expectation. There is a marginally decreasing benefit to the geometric mean for each market system added. That is, each new market system benefits the geometric mean to a lesser and lesser degree. Further, as you add each new market system, there is a greater and greater efficiency loss caused as a result of simultaneous rather than sequential outcomes. At some point, to add another market system may do more harm then good.

TIME REQUIRED TO REACH A SPECIFIED GOAL AND THE TROUBLE WITH FRACTIONAL *f*

Suppose we are given the arithmetic average HPR and the geometric average HPR for a given system. We can determine the standard deviation (SD) in HPRs from the formula for estimated geometric mean:

$$\mathrm{EGM} = \sqrt{\mathrm{AHPR}^2 - \mathrm{SD}^2}$$

where: AHPR = The arithmetic mean HPR.
SD = The population standard deviation in HPRs.

Therefore, we can estimate the SD as:

$$\mathrm{SD}^2 = \mathrm{AHPR}^2 - \mathrm{EGM}^2$$

Returning to our 2:1 coin-toss game, we have a mathematical expectation of \$.50, and an optimal *f* of betting \$1 for every \$4 in equity, which yields a geometric mean of 1.06066. We can use Equation (5.03) to determine our arithmetic average HPR:

$$\mathrm{AHPR} = 1 + (\mathrm{ME}/f\$) \tag{5.03}$$

where: AHPR = The arithmetic average HPR.
ME = The arithmetic mathematical expectation in units.
$f\$$ = The biggest loss/$-f$.
f = The optimal f (0 to 1).

Thus, we would have an arithmetic average HPR of:

$$\begin{aligned}\text{AHPR} &= 1 + (.5/(-1/-.25)) \\ &= 1 + (.5/4) \\ &= 1 + .125 \\ &= 1.125\end{aligned}$$

Now, since we have our AHPR and our EGM, we can employ Equation (5.04) to determine the estimated SD in the HPRs:

$$\begin{aligned}\text{SD}^2 &= \text{AHPR}^2 - \text{EGM}^2 \\ &= 1.125^2 - 1.06066^2 \\ &= 1.265625 - 1.124999636 \\ &= .140625364\end{aligned}$$

Thus, SD ^ 2, which is the variance in HPRs, is .140625364. Taking the square root of this yields an SD in these HPRs of .140625364 ^ (1/2) = .3750004853. You should note that this is the estimated SD because it uses the estimated geometric mean as input. It is probably not completely exact, but it is close enough for our purposes.

However, suppose we want to convert these values for the SD (or variance), arithmetic, and geometric mean HPRs to reflect trading at the fractional *f*. These conversions are now given:

$$\text{FAHPR} = (\text{AHPR} - 1) * \text{FRAC} + 1 \tag{5.04}$$

$$\text{FSD} = \text{SD} * \text{FRAC} \tag{5.05}$$

$$\text{FGHPR} = \sqrt{\text{FAHPR}^2 - \text{FSD}^2} \tag{5.06}$$

where: FRAC = The fraction of optimal *f* we are solving for.
AHPR = The arithmetic average HPR at the optimal *f*.
SD = The standard deviation in HPRs at the optimal *f*.
FAHPR = The arithmetic average HPR at the fractional *f*.
FSD = The standard deviation in HPRs at the fractional *f*.
FGHPR = The geometric average HPR at the fractional *f*.

For example, suppose we want to see what values we would have for FAHPR, FGHPR, and FSD at half the optimal *f* (FRAC = .5) in our 2:1 coin-toss game. Here, we know our AHPR is 1.125 and our SD is .3750004853. Thus:

$$\begin{aligned}\text{FAHPR} &= (\text{AHPR} - 1) * \text{FRAC} + 1 \\ &= (1.125 - 1) * .5 + 1 \\ &= .125 * .5 + 1 \\ &= .0625 + 1 \\ &= 1.0625\end{aligned}$$

$$\begin{aligned}\text{FSD} &= \text{SD} * \text{FRAC}\\ &= .3750004853 * .5\\ &= .1875002427\end{aligned}$$

$$\begin{aligned}\text{FGHPR} &= \sqrt{\text{FAHPR}^2 - \text{FSD}^2}\\ &= \sqrt{1.065^2 - .1875002427^2}\\ &= \sqrt{1.12890625 - .03515634101}\\ &= \sqrt{1.093749909}\\ &= 1.04582499\end{aligned}$$

Thus, for an optimal *f* of .25, or making one bet for every \$4 in equity, we have values of 1.125, 1.06066, and .3750004853 for the arithmetic average, geometric average, and SD of HPRs, respectively. Now we have solved for a fractional (.5) *f* of .125 or making one bet for every \$8 in our stake, yielding values of 1.0625, 1.04582499, and .1875002427 for the arithmetic average, geometric average, and SD of HPRs, respectively.

We can now take a look at what happens when we practice a fractional *f* strategy. We have already determined that under fractional *f* we will make geometrically less money than under optimal *f*. Further, we have determined that the drawdowns and variance in returns will be less with fractional *f*. What about time required to reach a specific goal?

We can quantify the expected number of trades required to reach a specific goal. This is not the same thing as the expected time required to reach a specific goal, but since our measurement is in trades we will use the two notions of time and trades elapsed interchangeably here:

$$T = \ln(\text{Goal}) / \ln(\text{Geometric Mean}) \tag{5.07}$$

where: T = The expected number of trades to reach a specific goal.
Goal = The goal in terms of a multiple on our starting stake, a TWR.
ln() = The natural logarithm function.

or:

$$T = \text{Log}_{\text{Geometric Mean}}\ \text{Goal}$$

$$\text{(i.e. The 'Log base Geoemetric Mean' of the Goal)} \tag{5.07a}$$

Returning to our 2:1 coin-toss example, at optimal *f* we have a geometric mean of 1.06066, and at half *f* this is 1.04582499. Now let's calculate the expected number of trades required to double our stake (goal = 2). At full *f*:

$$\begin{aligned} T &= \ln(2)/\ln(1.06066) \\ &= .6931471/.05889134 \\ &= 11.76993 \end{aligned}$$

Thus, at the full *f* amount in this 2:1 coin-toss game, we anticipate it will take us 11.76993 plays (trades) to double our stake.

Now, at the half *f* amount:

$$\begin{aligned} T &= \ln(2)/\ln(1.04582499) \\ &= .6931471/.04480602 \\ &= 15.46996 \end{aligned}$$

Thus, at the half *f* amount, we anticipate it will take us 15.46996 trades to double our stake. In other words, trading half *f* in this case will take us 31.44% longer to reach our goal.

Well, that doesn't sound too bad. By being more patient, allowing 31.44% longer to reach our goal, we eliminate our drawdown by half and our variance in the trades by half. Half *f* is a seemingly attractive way to go. The smaller the fraction of optimal *f* that you use, the smoother the equity curve, and hence the less time you can expect to be in the worst-case drawdown.

Now, let's look at it in another light. Suppose you open two accounts, one to trade the full *f* and one to trade the half *f*. After 12 plays, your full *f* account will have more than doubled to 2.02728259 (1.06066^{12}) times your starting stake. After 12 trades your half *f* account will have grown to 1.712017427 (1.04582499^{12}) times your starting stake. This half *f* account will double at 16 trades to a multiple of 2.048067384 (1.04582499^{16}) times your starting stake. So, by waiting about one-third longer, you have achieved the same goal as with full optimal *f*, only with half the commotion. However, by trade 16 the full *f* account is now at a multiple of 2.565777865 (1.06066^{16}) times your starting stake. Full *f* will continue to pull out and away. By trade 100, your half *f* account should be at a multiple of 88.28796546 times your starting stake, but the full *f* will be at a multiple of 361.093016!

So anyone who claims that the only thing you sacrifice with trading at a fractional versus full *f* is time required to reach a specific goal is completely correct. Yet time is what it's all about. We can put our money in Treasury bills and they will reach a specific goal in a certain time with an absolute minimum of drawdown and variance! Time truly is of the essence.

COMPARING TRADING SYSTEMS

We have seen that two trading systems can be compared on the basis of their geometric means at their respective optimal *f*s. Further, we can compare systems based on how high their optimal *f*s themselves are, with the higher optimal *f* being the riskier system. This is because the least the drawdown may have been is at least an *f* percent equity retracement. So, there are two basic measures for comparing systems, the geometric means at the optimal *f*s, with the higher geometric mean being the superior system, and the optimal *f*s themselves, with the lower optimal *f* being the superior system. Thus, rather than having a single, one-dimensional measure of system performance, we see that performance must be measured on a two-dimensional plane, one axis being the geometric mean, the other being the value for *f* itself. *The higher the geometric mean at the optimal* f, *the better the system. Also, the lower the optimal* f, *the better the system.*

Geometric mean does not imply anything regarding drawdown. That is, a higher geometric mean does not mean a higher (or lower) drawdown. The geometric mean pertains only to return. The optimal *f* is the measure of minimum expected historical drawdown as a percentage of equity retracement. A higher optimal *f* does not mean a higher (or lower) return. We can also use these benchmarks to compare a given system at a fractional *f* value and another given system at its full optimal *f* value.

Therefore, when looking at systems, you should look at them in terms of how high their geometric means are and what their optimal *f*s are. For example, suppose we have System A, which has a 1.05 geometric mean and an optimal *f* of .8. Also, we have System B, which has a geometric mean of 1.025 and an optimal *f* of .4. System A at the half *f* level will have the same minimum historical worst-case equity retracement (drawdown) of 40%, just as System B's at full *f*, but System A's geometric mean at half *f* will still be higher than System B's at the full *f* amount. Therefore, System A is superior to System B.

"Wait a minute," you say. "I thought the only thing that mattered was that we had a geometric mean greater than one, that the system need be only marginally profitable, that we can make all the money we want through money management!" That's still true. However, the rate at which you will make the money is still a function of the geometric mean at the *f* level you are employing. The expected variability will be a function of how high the *f* you are using is. So, although it's true that you *must* have a system with a geometric mean at the optimal *f* that is greater than one (i.e., a positive mathematical expectation) and that you can still make virtually an unlimited amount with such a system after enough trades, the rate of growth (the number of trades required to reach a specific goal) is dependent upon the geometric mean at the *f* value employed. The variability en route to that goal is also a function of the *f* value employed.

Yet these considerations, the degree of the geometric mean and the *f* employed, are *secondary* to the fact that you must have a positive mathematical expectation, although they are useful in comparing two systems or techniques that have positive mathematical expectations and an equal confidence of their working in the future.

TOO MUCH SENSITIVITY TO THE BIGGEST LOSS

A recurring criticism with the entire approach of optimal *f* is that it is too dependent on the biggest losing trade. This seems to be rather disturbing to many traders. They argue that the amount of contracts you put on today should not be so much a function of a single bad trade in the past.

Numerous different algorithms have been worked up by people to alleviate this apparent oversensitivity to the largest loss. Many of these algorithms work by adjusting the largest loss upward or downward to make the largest loss be a function of the current volatility in the market. The relationship seems to be a quadratic one. That is, the absolute value of the largest loss seems to get bigger at a faster rate than the volatility. (Volatility is usually defined by these practitioners as the average daily range of the last few weeks, or average absolute value of the daily net change of the last few weeks, or any of the other conventional measures of volatility.) However, this is not a deterministic relationship. That is, just because the volatility is X today does not mean that our largest loss *will* be X^Y. It simply means that it usually is *somewhere near* X^Y.

If we could determine in advance what the largest possible loss would be going into today, we could then have a much better handle on our money management.[2] Here again is a case where we must consider the worst-case

[2]This is where using options in a trading strategy is so useful. Either buying a put or call outright in opposition to the underlying position to limit the loss to the strike price of the options, or simply buying options outright in lieu of the underlying, gives you a floor, an absolute maximum loss. Knowing this is extremely handy from a money-management, particularly an optimal *f*, standpoint. Futher, if you know what your maximum possible loss is in advance (e.g., a day trade), then you can always determine what the *f* is in dollars perfectly for any trade by the relation dollars at risk per unit/optimal *f*. For example, suppose a day trader knew his optimal *f* was .4. His stop today, on a one-unit basis, is going to be $900. He will therefore optimally trade one unit for every $2,250 ($900/.4) in account equity.

scenario and build from there. The problem is that we do not know exactly what our largest loss can be going into today. An algorithm that can predict this is really not very useful to us because of the one time that it fails.

Consider, for instance, the possibility of an exogenous shock occurring in a market overnight. Suppose the volatility were quite low prior to this overnight shock, and the market then went locked-limit against you for the next few days. Or suppose that there were no price limits, and the market just opened an enormous amount against you the next day. These types of events are as old as commodity and stock trading itself. They can and do happen, *and they are not always telegraphed in advance* by increased volatility.

Generally, then, you are better off not to "shrink" your largest historical loss to reflect a current low-volatility marketplace. Furthermore, *there is the concrete possibility of experiencing a loss larger in the future than what was the historically largest loss.* There is no mandate that the largest loss seen in the past is the largest loss you can experience today. This is true regardless of the current volatility coming into today.

The problem is that, empirically, the *f* that has been optimal in the past is a function of the largest loss of the past. There's no getting around this. However, as you shall see when we get into the parametric techniques, you can budget for a greater loss in the future. In so doing, you will be prepared if the almost inevitable larger loss comes along. Rather than trying to adjust the largest loss to the current climate of a given market so that your empirical optimal *f* reflects the current climate, you will be much better off learning the parametric techniques.

The scenario planning techniques, which are a parametric technique, are a possible solution to this problem, and it can be applied whether we are deriving our optimal *f* empirically or, as we shall learn later, parametrically.

THE ARC SINE LAWS AND RANDOM WALKS

Now we turn the discussion toward drawdowns. First, however, we need to study a little bit of theory in the way of the first and second arc sine laws. These are principles that pertain to random walks. The stream of trade profits and losses (P&Ls) that you are dealing with may not be truly random. The degree to which the stream of P&Ls you are using differs from being purely random is the degree to which this discussion will not pertain to your stream of P&Ls. Generally, though, most streams of trade P&Ls are nearly random as determined by the runs test and the linear correlation coefficient (serial correlation).

Furthermore, not only do the arc sine laws assume that you know in advance the amount you can win or lose; they also assume that the amount you can win is equal to the amount you can lose, and that this is always a constant amount. In our discussion, we will assume that the amount you can win or lose is \$1 on each play. The arc sine laws also assume that you have a 50% chance of winning and a 50% chance of losing. Thus, the arc sine laws assume a game where the mathematical expectation is zero.

These caveats make for a game that is considerably different, and considerably simpler, than trading is. However, the first and second arc sine laws are exact for the game just described. To the degree that trading differs from the game just described, the arc sine laws do not apply. For the sake of learning the theory, however, we will not let these differences concern us for the moment.

Imagine a truly random sequence such as coin tossing[3] where we win one unit when we win and we lose one unit when we lose. If we were to plot out our equity curve over X tosses, we could refer to a specific point (X,Y), where X represented the Xth toss and Y our cumulative gain or loss as of that toss.

We define *positive territory* as anytime the equity curve is above the X axis or on the X axis when the previous point was above the X axis. Likewise, we define *negative territory* as anytime the equity curve is below the X axis or on the X axis when the previous point was below the X axis. We would expect the total number of points in positive territory to be close to the total number of points in negative territory. But this is not the case.

If you were to toss the coin N times, your probability (Prob) of spending K of the events in positive territory is:

$$\text{Prob} \sim 1/\pi * \sqrt{\text{K}} * \sqrt{(\text{N} - \text{K})} \tag{5.08}$$

The symbol $\sim$ means that both sides tend to equality in the limit. In this case, as either K or (N – K) approaches infinity, the two sides of the equation will tend toward equality.

[3]Although empirical tests show that coin tossing is not a truly random sequence due to slight imperfections in the coin used, we will assume here, and elsewhere in the text when referring to coin tossing, that we are tossing an ideal coin with exactly a .5 chance of landing heads or tails.

Thus, if we were to toss a coin 10 times (N = 10) we would have the following probabilities of being in positive territory for K of the tosses:

K	Probability[4]
0	.14795
1	.1061
2	.0796
3	.0695
4	.065
5	.0637
6	.065
7	.0695
8	.0796
9	.1061
10	.14795

You would expect to be in positive territory for 5 of the 10 tosses, yet that is the least likely outcome! In fact, the most likely outcome is that you will be in positive territory for all of the tosses or for none of them!

This principle is formally detailed in the *first arc sine law*, which states:

For a Fixed A (0 < A < 1) and as N approaches infinity, the probability that K/N spent on the positive side is < A tends to:

$$\text{Prob}\{(\text{K/N}) < \text{A}\} = 2/\pi * \sin^{-1}\sqrt{\text{A}} \tag{5.09}$$

Even with N as small as 20, you obtain a very close approximation for the probability.

Equation (5.09), the first arc sine law, tells us that with probability .1, we can expect to see 99.4% of the time spent on one side of the origin, and with probability .2, the equity curve will spend 97.6% of the time on the same side of the origin! With a probability of .5, we can expect the equity curve to spend in excess of 85.35% of the time on the same side of the origin. That is just how perverse the equity curve of a fair coin is!

[4]Note that since neither K nor N may equal 0 in Equation (5.08) (as you would then be dividing by 0), we can discern the probabilities corresponding to K = 0 and K = N by summing the probabilities from K = 1 to K = N – 1 and subtracting this sum from 1. Dividing this difference by 2 will give us the probabilities associated with K = 0 and K = N.

Now here is the *second arc sine law*, which also uses Equation (5.09) and hence has the same probabilities as the first arc sine law, but applies to an altogether different incident, the maximum or minimum of the equity curve. The second arc sine law states that the maximum (or minimum) point of an equity curve will most likely occur at the endpoints, and least likely at the center. The distribution is exactly the same as the amount of time spent on one side of the origin!

If you were to toss the coin N times, your probability of achieving the maximum (or minimum) at point K in the equity curve is also given by Equation (5.08):

$$\text{Prob} \sim 1/\pi * \sqrt{\text{K}} * \sqrt{(\text{N} - \text{K})}$$

Thus, if you were to toss a coin 10 times (N = 10), you would have the following probabilities of the maximum (or minimum) occurring on the Kth toss:

K	Probability
0	.14795
1	.1061
2	.0796
3	.0695
4	.065
5	.0637
6	.065
7	.0695
8	.0796
9	.1061
10	.14795

In a nutshell, the second arc sine law states that the maximum or minimum is most likely to occur near the endpoints of the equity curve and least likely to occur in the center.

TIME SPENT IN A DRAWDOWN

Recall the caveats involved with the arc sine laws. That is, the arc sine laws assume a 50% chance of winning and a 50% chance of losing. Further, they assume that you win or lose the exact same amounts and that the generating stream is purely random. Trading is considerably more complicated than this. Thus, the arc sine laws don't apply in a pure sense, but they do apply in spirit.

Consider that the arc sine laws worked on an arithmetic mathematical expectation of zero. Thus, with the first law, we can interpret the percentage of time on either side of the zero line as the percentage of time on either side of the arithmetic mathematical expectation. Likewise with the second law, where, rather than looking for an absolute maximum and minimum, we were looking for a maximum above the mathematical expectation and a minimum below it. The minimum below the mathematical expectation could be greater than the maximum above it if the minimum happened later and the arithmetic mathematical expectation was a rising line (as in trading) rather than a horizontal line at zero.

However we can interpret the spirit of the arc sine laws as applying to trading in the following ways. First, each trade, regardless of the amount won or lost, must be considered as winning one unit or losing one unit respectively. Thus, we now therefore have a line whose slope is the ratio of the difference between the number of wins and losses, and the sum of the number of wins and number of losses, rather than the horizontal line whose slope is zero in the arc sine laws.

For example, suppose I had four trades, three of which were winning trades. The slope of my line therefore equals $(3-1)/(3+1) = 2/4 = .5$. This is our slope and our mathematical expectation (given that all wins are figured as $+1$, all losses as -1).

We can interpret the first arc sine law as stating that we should expect to be on one side of the mathematical expectation line for far more trades than we spend on the other side of the mathematical expectation line. Regarding the second arc sine law, we should expect the maximum deviations from the mathematical expectation line, either above or below it, as being most likely to occur near the beginning or the end of the equity curve graph and least likely near the center of it.

THE ESTIMATED GEOMETRIC MEAN (OR HOW THE DISPERSION OF OUTCOMES AFFECTS GEOMETRIC GROWTH)

This discussion will use a gambling illustration for the sake of simplicity. Let's consider two systems: System A, which wins 10% of the time and has a twenty-eight-to-one win/loss ratio, and System B, which wins 70% of the time and has a one-to-one ratio. Our mathematical expectation, per unit bet, for A is 1.9 and for B is .4. Therefore, we can say that for every unit bet, System A will return, on average, 4.75 times as much as System B. But let's examine this under fixed fractional trading. We can find our optimal fs by dividing the mathematical expectations by the win/loss ratios [per

Equation (4.05)]. This gives us an optimal f of .0678 for A and .4 for B. The geometric means for each system at their optimal f levels are then:

$$A = 1.044176755$$
$$B = 1.0857629$$

System	% Wins	Win:Loss	ME	*f*	Geomean
A	.1	28:1	1.9	.0678	1.0441768
B	.7	1:1	.4	.4	1.0857629

As you can see, System B, although less than one-fourth the mathematical expectation of A, makes almost twice as much per bet (returning 8.57629% of your entire stake per bet, on average, when reinvesting at the optimal f levels) as does A (returning 4.4176755% of your entire stake per bet, on average, when reinvesting at the optimal f levels).

Now, assuming a 50% drawdown on equity will require a 100% gain to recoup, then:

> 1.044177 to the power of x is equal to 2.0 at approximately x equals 16.5, or more than 16 trades to recoup from a 50% drawdown for System A. Contrast this to System B, where 1.0857629 to the power of x is equal to 2.0 at approximately x equals 9, or nine trades for System B to recoup from a 50% drawdown.

What's going on here? Is this because System B has a higher percentage of winning trades? The reason B is outperforming A has to do with the dispersion of outcomes and its effect on the growth function. Most people have the mistaken impression that the growth function, the TWR, is:

$$\text{TWR} = (1 + R)^T$$

where: R = Interest rate per period, e.g., 7% = .07.
T = Number of periods.

Since $1 + R$ is the same thing as an HPR, we can say that most people have the mistaken impression that the growth function,[5] the TWR, is:

$$\text{TWR} = \text{HPR}^T$$

[5]Many people mistakenly use the arithmetic average HPR in the equation for HPR^T. As is demonstrated here, this will not give the true TWR after T plays. What you must use is the geometric average HPR, rather than the arithmetic in HPR^T. This will give you the true TWR. If the standard deviation in HPRs is 0, then the arithmetic average HPR and the geometric average HPR are equivalent, and it matters not which you use, arithmetic or geometric average HPR, in such a case.

This function is true only when the return (i.e., the HPR) is constant, which is not the case in trading.

The real growth function in trading (or any event where the HPR is not constant) is the multiplicative product of the HPRs. Assume we are trading coffee, and our optimal f is one contract for every \$21,000 in equity, and we have two trades, a loss of \$210 and a gain of \$210, for HPRs of .99 and 1.01, respectively. In this example, our TWR would be:

$$\begin{aligned}\text{TWR} &= 1.01 * .99\\ &= .9999\end{aligned}$$

An insight can be gained by using the estimated geometric mean (EGM), which very closely approximates the geometric mean:

$$G = \sqrt{A^2 - S^2}$$

or:

$$G = \sqrt{A^2 - V}$$

where: G = geometric mean HPR
A = arithmetic mean HPR
S = standard deviation in HPRs
V = variance in HPRs

Now we take Equations (4.18) and (3.04) to the power of n to estimate the TWR. This will very closely approximate the *multiplicative* growth function, the actual TWR, of Equation (4.17):

$$\text{TWR} = (\sqrt{A^2 - S^2})^T \tag{5.10}$$

where: T = Number of periods.
A = Arithmetic mean HPR.
S = Population standard deviation in HPRs.

The insight gained is that we can see, mathematically, the trade-off between an increase in the arithmetic average trade (the HPR) versus the dispersion in the HPRs (the standard deviations or the variance), hence the reason that the 70% one-to-one system did better than the 10% twenty-eight-to-one system.

Our goal should be to maximize the coefficient of this function, to maximize Equation (3.04): Expressed literally, to maximize *the square root of the quantity HPR squared minus the variance in HPRs.*

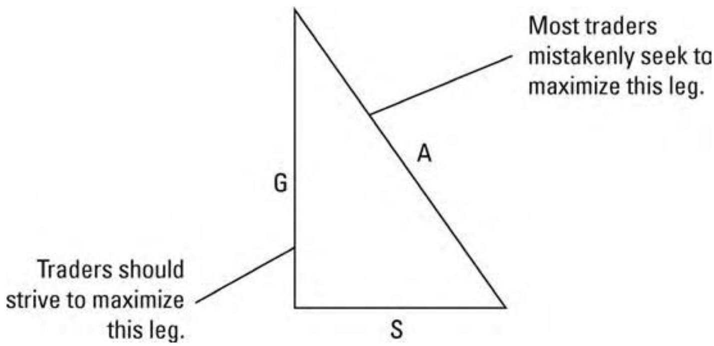

FIGURE 5.2 Pythagorean Theorem in money management

The exponent of the estimated TWR, T, will take care of itself. That is to say that increasing T is not a problem, as we can increase the number of markets we are following, trading more short-term types of systems, and so on.

We can rewrite Equation [3.04] to appear as:

$$A^2 = G^2 + S^2$$

This brings us to the point where we can envision exactly what the relationships are. Notice that this equation is the familiar Pythagorean Theorem: The hypotenuse of a right-angle triangle squared equals the sum of the squares of its sides (Figure 5.2). But here, the hypotenuse is A, and we want to maximize one of the legs, G.

In maximizing G, any increase in S will require an increase in A to offset. When S equals zero, then A equals G, thus conforming to the misconstrued growth function $\text{TWR} = (1 + R)^T$.

So, in terms of their relative effect on G, we can state that an increase in A is equal to a decrease of the same amount in S, and vice versa. Thus, any amount by which the dispersion in trades is reduced (in terms of reducing the standard deviation) is equivalent to an increase in the arithmetic average HPR. This is true regardless of whether or not you are trading at optimal *f*!

If a trader is trading on a fixed fractional basis, then he wants to maximize G, not necessarily A. In maximizing G, the trader should realize that the standard deviation, S, affects G in directly the same proportion as does A, per the Pythagorean Theorem! Thus, when the trader reduces the standard deviation (S) of his trades, it is equivalent to an equal increase in the arithmetic average HPR (A), and vice versa!

THE FUNDAMENTAL EQUATION OF TRADING

We can glean a lot more than just how trimming the size of our losses, or reducing our dispersion in trades, improves our bottom line. Return now to Equation (5.10), the estimated TWR. Since $(X^Y)^Z = \mathrm{X}^{(Y*Z)}$, we can further simplify the exponents in the equation, thus simplifying Equation (5.10) to:

$$\mathrm{TWR} = (A^2 - S^2)^{T/2} \tag{5.10a}$$

This last equation, the simplification for the estimated TWR, we will call the fundamental equation for trading, since it describes how the different factors, A, S, and T, affect our bottom line in trading.

There are a few things that are readily apparent. The first of these is that if A is less than or equal to one, then regardless of the other two variables, S and T, our result can be no greater than one. If A is less than one, then as T approaches infinity, A approaches zero. This means that if A is less than or equal to one (mathematical expectation less than or equal to zero since mathematical expectation $= A - 1$), we do not stand a chance at making profits. In fact, if A is less than one, it is simply a matter of time until we go broke.

Provided that A is greater than one, we can see that increasing T increases our total profits. For each increase of one trade, the coefficient is further multiplied by its square root.

Each time we can increase T by one, we increase our TWR by a factor equivalent to the square root of the coefficient (which is the geometric mean). Thus, each time a trade occurs or an HPR elapses, each time T is increased by one, the coefficient is multiplied by the geometric mean.

An important point to note about the fundamental trading equation is that it shows that if you reduce your standard deviation to a greater extent than you reduce your arithmetic average HPR, you are better off. It stands to reason, therefore, that cutting your losses short, if possible, benefits you. But the equation demonstrates that at some point you no longer benefit by cutting your losses short. That is the point where you would be getting out of too many trades with a small loss that later would have turned profitable, thus reducing your A to a greater extent than your S.

Along these same lines, reducing big winning trades can help your program if it reduces your S greater than it reduces your A. This can be accomplished, in many cases, by incorporating options into your trading program. Having an option position that goes against your position in the underlying (either by buying long an option or writing an option) can possibly help.

As you can see, the fundamental trading equation can be utilized to dictate many changes in our trading. These changes may be in the way of tightening (or loosening) our stops, setting targets, and the like. These changes are the result of inefficiencies in the way we are carrying out our trading, as well as inefficiencies in our trading program or methodology.

WHY IS *f* OPTIMAL?

To see that f is optimal in the sense of maximizing wealth:

$$\text{since } G = \left(\prod_{i=1}^{T} \text{HPR}_i\right)^{1/T}$$

$$\text{and} \qquad \left(\prod_{i=1}^{T} \text{HPR}_i\right)^{1/T} = \exp\left(\frac{\sum_{i=1}^{T} \text{In}\,(\text{HPR}_i)}{T}\right)$$

Then, if one acts to maximize the geometric mean at every holding period, if the trial is sufficiently long, by applying either the weaker law of large numbers or the Central Limit Theorem to the sum of *independent* variables (i.e., the numerator on the right side of this equation), almost certainly higher terminal wealth will result than from using any other decision rule.

Furthermore, we can also apply Rolle's Theorem to the problem of the proof of fs optimality. Recall that we are defining *optimal* here as meaning that which will result in the greatest geometric growth as the number of trials increases. The TWR is the measure of average geometric growth; thus, we wish to prove that there is a value for f that results in the greatest TWR.

Rolle's Theorem states that if a *continuous* function crosses a line parallel to the X-axis at two points, a and b, and the function is continuous throughout the interval a,b, then there exists at least one point in the interval where the first derivative equals zero (i.e., at least one relative extremum).

Given that all functions with a positive arithmetic mathematical expectation cross the X-axis twice[6] (the X being the f axis), at $f = 0$ and at that point to the right where f results in computed HPRs where the variance in those HPRs exceeds the difference of the arithmetic mean of those HPRs minus one, we have our a,b interval on X, respectively. Furthermore, the

[6]Actually, at $f = 0$, the TWR $= 0$, and thus we cannot say that it crosses 0 to the upside here. Instead, we can say that at an f value which is an infinitesimally small amount beyond 0, the TWR crosses a line an infinitesimally small amount above 0. Likewise to the right but in reverse, the line, the f curve, the TWR, crosses this line which is an infinitesimally small amount above the X-axis as it comes back down to the X-axis.

first derivative of the fundamental equation of trading (i.e., the estimated TWR) is continuous for all f within the interval, since f results in AHPRs and variances in those HPRs, within the interval, which are differentiable in the function in that interval; thus, the function, the estimated TWR, is continuous within the interval. Per Rolle's Theorem, it must, therefore, have at least one relative extremum in the interval, and since the interval is positive, that is, above the X-axis, the interval must contain at least one maximum.

In fact, there can be only one maximum in the interval given that the change in the geometric mean HPR (a transformation of the TWR, given that the geometric mean HPR is the Tth root of the TWR) is a direct function of the change in the AHPR and the variance, both of which vary in *opposite directions to each other as* f *varies*, per the Pythagorean theorem. This guarantees that there can be only one peak. Thus, there must be a peak in the interval, and there can be only one peak. There is an f that is optimal at only one value for f, where the first derivative of the TWR with respect to f equals zero.

Let us go back to Equation (4.07). Now, we again consider our two-to-one coin toss. There are two trades, two possible scenarios. If we take the first derivative of (4.07) with respect to f, we obtain:

$$\begin{aligned}\frac{d\mathrm{TWR}}{df} &= \left(\left(1+f^{*}\left(\frac{-\mathrm{trade}_1}{\text{biggest loss}}\right)\right)*\left(\frac{-\mathrm{trade}_2}{\text{biggest loss}}\right)\right)\\ &\quad+\left(\left(\frac{-\mathrm{trade}_1}{\text{biggest loss}}\right)*\left(1+f*\left(\frac{-\mathrm{trade}_2}{\text{biggest loss}}\right)\right)\right) \qquad (5.11)\end{aligned}$$

If there were more than two trades, the same basic form could be used, only it would grow monstrously large in short order, so we'll use only two trades for the sake of simplicity. Thus, for the sequence +2,−1 at $f = .25$:

$$\frac{d\mathrm{TWR}}{df} = \left(\left(1+.25*\left(\frac{-2}{-1}\right)\right)*\left(\frac{-1}{-1}\right)\right)+\left(\left(\frac{-2}{-1}\right)*\left(1+.25*\left(\frac{-1}{-1}\right)\right)\right)$$

$$\frac{d\mathrm{TWR}}{df} = ((1+.25*2)\;*-1)+(2*(1+.25*-1))$$

$$\frac{d\mathrm{TWR}}{df} = ((1+.5)*-1)+(2*(1-.25))$$

$$\frac{d\mathrm{TWR}}{df} = (1.5*-1)+(2*.75)$$

$$\frac{d\mathrm{TWR}}{df} = -1.5+1.5 = 0$$

And we see that the function peaks at .25, where the slope of the tangent is zero, exactly at the optimal f, and no other local extremum can exist because of the restriction caused by the Pythagorean Theorem.

Lastly, we will see that optimal f is indifferent to T. We can take the first derivative of the estimated TWR, Equation (5.10a) with respect to T as:

$$\frac{d\text{TWR}}{dT} = \left(A^2 - S^2\right)^{T/2} * \ln\left(A^2 - S^2\right) \tag{5.12}$$

Since $\ln(1) = 0$, then if $A^2 - S^2 = 1$, that is, $A^2 - 1 = S^2$ (or variance), the function peaks out and the single optimal maximum TWR is found with respect to f. Notice, though, that both A, the arithmetic average HPR, and S, the standard deviation in those HPRs, are not functions of T. Instead, they are indifferent to T; thus, (5.10a) is indifferent to T at the optimal f. The f that is optimal in the sense of maximizing the estimated TWR will always be the same value regardless of T.