An easy-to-use measure will allow you to quickly and reliably evaluate the efficiency that your trading system achieves. Static evaluation ignores sequence and distribution of trades over time, but a new dynamic risk-monitoring indicator helps take account of large consecutive strings of losses properly.

How best to evaluate system performance

BY DAVID SEPIASHVILI

Learning how to judge and compare trading systems is at least as important as actually building them. One problem is a choice of an appropriate performance measurement. Using the right criteria is an important cornerstone not only in an adequate assessment of trading systems but in finding the most promising avenues for their further improvement or designing new trading strategies.

Here we will cover a simple measure for performance evaluation called the Efficiency Factor. This measure fully integrates all the essential ways of assessing a trading system's profitability while remaining transparent in the physical sense and is easy to calculate by hand.

Efficiency, a common sense energy concept and a closely related concept of productivity, is used in the widest fields of applications. It is generally measured by the ratio of output to input of any system. An increase in the ratio indicates an increase in efficiency, and a decrease in the ratio signals a decline in efficiency. A trading system's performance is not an exception. Efficiency factor — the ratio of the effective or useful output to the total input — must find its place in the overall evaluation and comparison of the strategy performance reports as an effective and reliable measurement of a trading system's worth.

The efficiency factor in trading terms is the ratio of net profit to gross profit. The formula of the efficiency factor is intuitive and simple to use:

EF = NP / GP

However, it offers quite a reliable measure of a system's performance, its particular strength being the elements of its intrinsic structure and their inherent correlation.

EFFECTIVE AND RELIABLE

The most obvious characteristics of trading strategy performance, given in testing summaries, are total number of trades, including winning trades, gross profit and gross loss. Based on these primary measures, two essential performance measurements — win probability and win/loss ratio — are derived. Frequently described as "percentage profitable," win probability is calculated by dividing the number of winning trades n(w) by total number of trades n, and win/loss ratio is found by dividing "average win per winning trade" by "average loss per losing trade."

Unfortunately, considered in isolation, these characteristics cannot give a full and comprehensive picture of a trading system's performance. Taken individually, they are not indicative of a system's true overall performance. They can even be in opposite directions to each other; a low win/loss ratio may go with a high win probability and vice versa. The efficiency factor combines them into one integrated measurement, which can be a helpful characteristic of likely success in further trading.

Let us get an insight into the EF formula. It also can be represented as:

$$EF = (GP - GL) / GP$$
$$= 1 - (p(l) / p(w)) * 1 / (W / L)$$

Where,

GP = n(w) * W = gross profit;
GL = n(l) * L = gross loss;
p(w) = n(w) / n = win probability;
p(l) = n(l) / n = loss probability;
W = average win per winning trade;
L - average loss per losing trade;
W/L = win/loss ratio.

Because the win probability and the loss probability are related by the equation...

$$p(w) + p(l) = 1$$

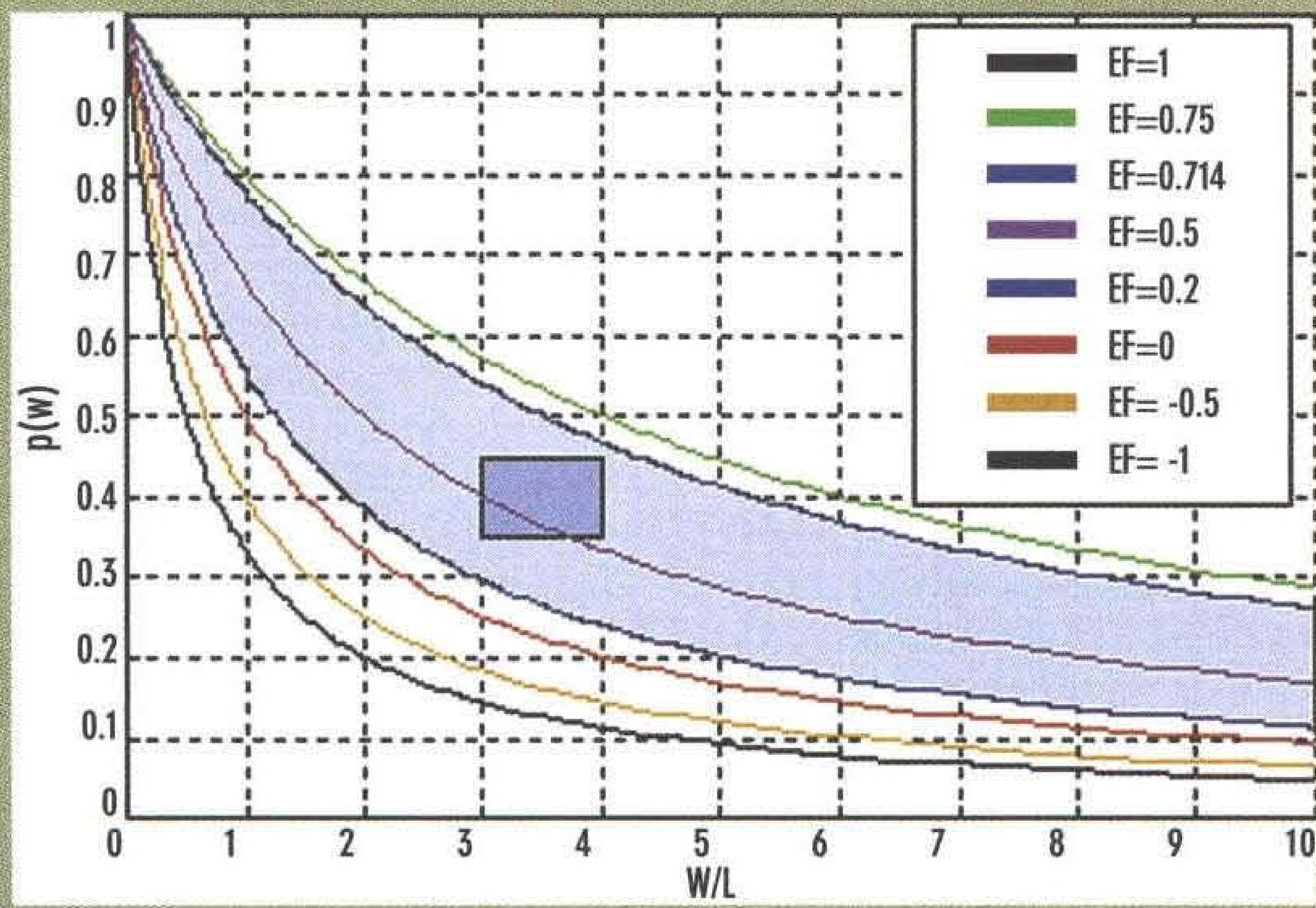
...our formula can be rewritten as:

$$EF = 1 - ((1 - p(w)) / p(w)) * (1 / (W / L))$$

Now, consider an example of EF estimation. Many successful traders believe that achieving values of win probabilities from 35% to 45% and of win/loss ratios from 3.00 to 4.00 are considered to be good for a trading system. Efficiency Factor for the given values, according to the final variation of our equation varies in the interval of 0.38 to 0.69. The higher the efficiency factor, the better the performance and the higher success in trading.

EFFICIENCY NOMOGRAM

This demonstrates how a set of achievable pairs of win probability and win/loss ratio determines the values of the Efficiency Factor.



TRUE COLORS

Compared side to side, you can see how the EF value is a valid representation of a trading system's overall worth.

Systems	n(w)	n	(GP- GL),%	NP,%	p(w)	W/L	EF	NP/trade,%
(7-15)	7	20	44.4-32.9	11.5	0.35	2.51	0.26	0.58
(10-20)	6	14	48.6-29.8	18.8	0.43	2.17	0.39	1.34
(12-25)	5	9	51.8-14.9	36.9	0.56	2.78	0.71	4.10
(10-30)	5	10	54.4-16.8	37.6	0.50	3.23	0.69	3.76
(10-40)	3	7	56.0-12.8	43.2	0.43	5.83	0.77	6.17

WHAT'S THE ROLE OF THE TWO MEASURES?

Which of the two components, win probability or win/loss ratio, plays a greater role in improving the efficiency factor of the system?

The last version of our formula above lets us analyze the impact of these two components on the overall performance and reveal their correlation. Consider an example. The lowest value for successful trading in the example (EF = 0.38) can be improved in two ways — either by increasing win/loss ratio or by enhancing win probability.

By changing win/loss ratio from 3.0 to 4.0 (an increase of 33.33%), we improve EF to 0.536. By changing win probability from 0.35 to 0.467 (an increase of 33.33%), we get an EF of 0.62. Based on that exercise, we see that increasing the values of the components by the same percentage 33%, they result in a different increase in the EF value.

Efficiency Factor is more sensitive to changes in win probability than to changes in win/loss ratio. Therefore, strategies directed at improving Efficiency Factor by reducing a number of losing trades through recognizing the circumstances that produce them are more efficient than strategies directed at reducing the magnitudes of all losses, using, for example, stop-loss orders.

CORRELATION IN A PAIR

Solving the equation for the variable p(w), we have...

$$p(w) = 1 / (1 + (1 - EF) * (W/L))$$

...for any distribution on EF.

As you can see from this formula, p(w) and (W/L) are not arbitrary — they are connected via EF. That's why they cannot be considered isolated.

What we will term the "Sepiashvili Nomogram" in "Efficiency nomogram" (above) gives a graphic representation of numerical interrelations between the win probability and the win/loss ratio.

Win probability values lie on the Y-axis and win/loss ratio values are displayed along the X-axis. The nomogram illustrates a set of achievable p(w) and (W/L) pairs, which lie along the isocurves of various fixed values of EF serving as efficiency boundaries.

ISOCURVES TO HELP

To determine the upper boundary of win probability let's introduce a hypothetical risk-free trading system, providing an EF = 1. The case corresponds to a point p(w) = 1 on the Y-axis and an arbitrary (W/L). It is obvious that the case is of theoretical interest only, as the real values of win probability always lie well below this boundary.

To determine the lower boundary, consider that the trader has lost all his initial trading capital: GL = 2 GP, EF = (-1). To continue trading below this level he must invest additional funds.

The isocurve corresponding to the case of zero efficiency (EF = 0) separates the regions of positive and negative returns. The more the distance of the p(w), (W/L) pair from this curve in the positive direction, the higher the performance of the trading system.

The middle isocurve between EF = 0 and EF = (-1) representing the case of 50% loss of the initial trading capital, can be considered as a benchmark of "ruin." That is the level under which you have to be a very courageous trader to keep at it.

In the region of positive returns, there are two more curves charted: EF = 0.5 and EF = 0.75. They can be interpreted as benchmarks of successful trading and successful investment, respectively.

The nomogram allows us to step through the isocurves mentioned above. It shows how different data sets would be located, by highlighting the appropriate isocurves on the nomogram and displaying the region of the most favorable values, as we progress through it.

The region of "real," actual efficiency, corresponding to underlying market retracements from 33.3% to 66.6%, is shaded on the nomogram. The upper boundary of the region (EF = 0.714) correlates with the best scenario and the lower boundary (EF = 0.2) with the worst scenario. The isocurve EF = 0.5 corresponds to a 50% retracement. Note that there is no blue bird on the market, even with the most accurate timing, you cannot exceed the upper boundary of the region on average.

This set of isocurves expedites search and location of EF values. It helps you quickly evaluate the EF your trading system achieves. The nomogram also lets you analyze the performance of the different p(w) and (W/L) under various market conditions conveniently and quickly, anytime.

SHORT-TERM VS. LONG-TERM

As you can see from the nomogram, in the interval of small win/loss ratios (for example (W/L) < 4.0), the win probability decays rapidly. At large values of win/loss ratios (for example (W/L) > 4), the slope of the curves decreases. That confirms our earlier findings about the efficiency factor's sensitivity. Higher values of efficiency factor are achieved mainly due to higher win probability and that is the advantage of long-term investment.

As our studies show, in strong-trending markets, short-term strategies generate three times more trades, while obtaining net profits only 1.3 times higher, even if your timing is perfect. If you take into account transaction fees, especially when you have limited initial trading capital, and lower win probability of short-term trading, the advantage of long-term investment is obvious.

Short-term trading strategies can be efficient only for moderately trending or wide price-ranging markets, in which returns of short-term strategies theoretically exceed returns of long-term investment by 1.5 to 1.8 times, while the number of trades is still three times greater.

NET PROFIT PER TRADE

A performance summary can include the value of net profit per trade. The expected value of net profit per trade is described by the equation:

$$NP / n = (GP / n) - (GL / n) = p(w) * W - p(l) * L$$

For correct calculation of the measure, note its distinction from the X-factor. (See "The X-factor: Synthesizing trading system reports," November 2001). Our measure reduces, in its essence, to the formula:

$$(NP / n) * (1 / L) = p(w) * (W/L) - p(l)$$

CROSSOVER CHALLENGE

For Exxon (XOM) directional divergences between signals generated by a 10-day and 40-day length exponential moving average crossover and DDI readings point to unpromising trading opportunities. For Broadvision (BVSN), the DDI and CPI indicators complement each other to show how consecutive losses generated by the seven-day and 15-day exponential moving average crossover system kills the high gross profit, reducing the resulting profit by more than three times.

Systems	n(w)	n	(GP-GL),%	NP,%	p(w)	W/L	EF	NP/trade
Exxon Mobil Corp. (10-40) (Oct. 1, 2001 -July 1, 2003)	2	11	14.2-42.1	(-27.9)	0.18	1.52	(-1.96)	(-2.54)
Broadvision Inc. (7-15) (July 1, 2002 -July 1, 2004)	4	21	173.4-109.5	63.9	0.19	6.73	0.37	3.04

If we assume, for example, that W = 4 and L = 2, then W/L = 2. To get an adequate estimate, take care to substitute 4 and 2 into our formula, and not their ratio 2:1, which is used in the X-factor's calculation.

VALIDITY OF EF

By way of illustration of the Efficiency Factor's usefulness and validity, "True colors" (page 37) presents examples of historical performance of one simple but popular trading model based on two exponential moving average crossovers for Russell 2000 index iShares (IWM) during Jan. 1, 2002, through July 1, 2004.

As you can see from the examples, the Efficiency Factor provides an objective basis for the trading systems' comparative analysis. All w(p) and (W/L) pairs lie in the region of positive returns. Except for the (7-15) model, they demonstrate sufficient levels of performance: p(w) changes in the interval (0.43 - 0.56), (W/L) changes in the interval (2.17 - 5.83), EF changes from 0.39 to 0.77. But only the comparison of total number of trades, Efficiency Factor and net profit per trade shows that the trading system (10-40) is the best.

What is also attractive in EF measurement is that it indicates the deviation from the normalized value of 1.00, and that makes it clearly defined and easily comparable. While past performance is no indication of future returns, you still can make some useful observations from information of this kind and select a model most suitable for your target profit and risk tolerance.

Now place the values p(w), (W/L) on the Efficiency Nomogram and compare them against the realistic set of isocurves. Are you inside or outside the shaded zone? It also is interesting to compare the results against average values achieved by successful traders in trading practice, which are marked on the nomogram by a rectangular. As you can see, three from five combinations of the two moving average crossover system in our demo example produce results corresponding to the upper boundary of efficiency that practically can be achieved. But note that it was an example of the best scenario. The interplay of two moving averages creates an easy-to-use trading rule, but, as we will see, it is often out of phase during sideways markets.

MAXIMUM CONSECUTIVE DRAWDOWNS

A system's performance evaluation is not complete if risk is not considered. The maximum drawdown is the most widespread risk measure in performance summaries among traders.

The maximum drawdown can be loosely defined as the largest paper loss of a single trade that has occurred during the testing period. Some analysts consider the concept evasive and not quite reliable.

It makes good sense to think of "maximum drawdown" also in a broader meaning as a maximum cumulative drawdown of consecutive losing trades — a risky phenomenon that should be accounted for by all traders.

RISK MONITORING INDICATORS

Static evaluation has some inherent limitations, primary of which is that trading performance and risk characteristics are factored only in the final results, and ignored in sequence over time. To control risk dynamically, some specific indicators are necessary.

To measure maximum losses/profits dynamically, and control the running values of cumulative losses/profits and the effect of consecutive trading losses that many traders find unacceptable to sustain, a new cumulative profit indicator (CPI) is presented here. Note that gains are withdrawn or losses are added at the start of a trade so that the same amount of money is invested into each trade.

Along with other functions, the indicator reflects consecutive trading losses produced by a system. Three running curves permit control of risk/return with regard to profit targets and acceptable risk tolerance levels for both single trades and a string of trades. A level of (-50%) is a benchmark of risk of ruin, under which a trading model "self-destructs."

The indicator combines well with our two other risk-monitoring tools — Directional Divergence and Profit-Loss Control Indicator (DDI) and Trend Quality Indicator (Q-Indicator).

We can illustrate the offered approach and risk-monitoring indicators with examples of a two moving averages crossover system. In "Crossover challenge" (page 38), you can see the performance results of the trading rule applied to Exxon Mobil Corp. and Broadvision Inc. You also can see how you could improve them by using our risk-monitoring tools.

These results are based on a 10-day and 40-day exponential moving average crossover system in Exxon and by a seven-day and 15-day exponential moving average crossover system in Broadvision during a trading range market. A running profit curve, a running loss curve, and a running net profit curve, along with a benchmark of risk of ruin (50%) were used.

Used together, the DDI and the CPI help control risk/return with regard to target profits and acceptable risk tolerance levels for both single trades and a series of trades. The Q-Indicator keeps you from responding to every little zigzag when the prices lie in the narrow ranges, yet points to the developing trends. Note how the indicator's readings in "dead" zones, or in zones of weak trends coincide with unprofitable periods marked by the DDI's consecutive losses. FM

David Sepiashvili is a doctorate candidate in signal processing at Carnegie Mellon University in Pittsburgh. Financial time series analysis is one of the applications of his scientific efforts. He is president of Alticom Inc., which has developed a number of solutions for stock market technical analysis (www.alticom.com). Reach him at das@alticom.com.