

3

A Perspective on Quantitative Finance: Models for Beating the Market

Ed Thorp

Quantitative Finance Review 2003

This is a perspective on quantitative finance from my point of view, a 45-year effort to build mathematical models for “beating markets”, by which I mean achieving risk-adjusted excess returns.

I’d like to illustrate with models I’ve developed, starting with a relatively simple example, the widely played casino game of blackjack or twenty-one. What does blackjack have to do with finance? A lot more than I first thought, as we’ll see.

Blackjack

When I first learned of the game in 1958, I was a new PhD in a part of mathematics known as functional analysis. I had never gambled in a casino. I avoided negative expectation games. I knew of the various proofs that it was not possible to gain an edge in virtually all the standard casino gambling games. But an article by four young mathematicians presented a “basic” strategy for blackjack that allegedly cut the House edge to a mere 0.6%. The authors had developed their strategy for a complete randomly shuffled deck.

But in the game as played, successive rounds are dealt from a more and more depleted pack of cards. To me, this “sampling without replacement”, or “dependence of trials”, meant that the usual proofs that you couldn’t beat the game did not apply. The game might be beatable. I

Contact address: Edward O. Thorp & Associates, 610 Newport Center Drive, Suite 1240, Newport Beach, CA 92660, USA.
E-mail: EOThorp@ix.netcom.com

realized that the player's expectations would fluctuate, under best strategy, depending on which depleted pack was being used. Would the fluctuations be enough to give favorable betting opportunities? The domain of the expectation function for one deck had more than 33 million points in a space with 10 independent variables, corresponding to the 10 different card values in blackjack (for eight decks it goes up to 6×10^{15}).

Voila! There "must be" whole continents of positive expectation. Now to find them. The paper I read had found the strategy and expectation for only one of these 33 million points. That was only an approximation with a smallish but poorly known error term, and it took 12 years on desk calculators. And each such strategy had to address several mathematically distinct decisions at the table, starting with 550 different combinations of the dealer's up card and the player's initial two cards. Nonetheless, I had taken the first step towards building a model: the key idea or "inspiration" – the domain of the visionaries.

The next step is to develop and refine the idea via quantitative and technical work so that it can be used in the real world. The brute force method would be to compute the basic strategy and expectation for each of the 33 million mathematically distinct subsets of cards and assemble a 33 million-page "book". Fortunately, using linear methods well known to me from functional analysis, I was able to build a simplified approximate model for much of the 10-dimensional expectation surface. My methods reduced the problem from 400 million years on a desk calculator to a few hundred years – still too long. However, I was fortunate to have moved to MIT at the beginning of the computer revolution, where as a faculty member I had access to one of the early high-speed mainframe computers, the IBM 704.

In short sessions over several months the computer generated the results I needed in about 2 hours of actual computer time. I was then able to condense all the information into a simplified card counting scheme and corresponding strategy tables that were only moderately more complex than the original "no memory" basic strategy for a complete deck. You can see them in my book *Beat the Dealer*. This work was the second step, the development of the idea into a model that can be tested in the real world – the domain of the quants.

When I announced my results to the mathematical community, they were surprised to see, for the first time, contrary to long-held beliefs, a winning mathematical system for a major casino gambling game. They generally understood and approved. Not so, universally, however. Several casinos said they loved system players and would send a cab to the airport for me. The *Washington Post* greeted me with an editorial in January, 1961, saying:

"We hear there's a mathematician coming to town who claims he can beat blackjack. It reminds us of an ad we saw: 'Sure fire weed killer – send in \$1'. Back came a postcard saying, 'Grab by the roots and pull like hell'."

So I took the third and last step in building a successful model, real world verification – the domain of the entrepreneurs. In 20 hours of full-scale betting on a first casino trip in 1961 I won \$11,000, using a \$10,000 bankroll. The midpoint of my forecast was a win of \$10,000. To convert to today's dollars, multiply by about seven.

To recap, the three steps for a successful market-beating model are (1) idea, (2) development, and (3) successful real world implementation. The relevant skills are (1) visionary, (2) quantitative, and (3) entrepreneurial.

Early on, I assessed the blackjack idea as worth a million dollars to me, meaning that if I chose to focus my time and energy on exploiting it I could personally extract a million after tax dollars, net of costs, within a few years.

The relevance to finance proved to be considerable. First, it showed that the “gambling market” was, in finance theory language, inefficient (i.e. beatable); if that was so, why not the more complex financial markets? Second, it had a significant worldwide impact on the financial results of casinos. Although it did create a plague of hundreds and eventually thousands of new experts who extracted hundreds of millions from the casinos over the years, it also created a windfall of hundreds of thousands of hopefuls who, although improved, still didn’t play well enough to have an edge. Blackjack and the revenues from it surged. Third, it popularized a method that could be used to manage risk more generally in the investment world, namely the Kelly criterion, a.k.a. the capital growth criterion, a.k.a. fixed fraction betting, a.k.a. maximizing expected logarithmic utility.

Now a word on what I learned from this about risk control. The field trip to Nevada was bankrolled by two wealthy businessmen/gamblers, whom I chose from the many who sought me out.

- (a) They proposed a \$100K bankroll.
- (b) I reduced it to \$10K as a personal safety measure.
- (c) I began play by limiting my betting to \$1–10 for the first 8 hours – to verify all was as expected (possible cheating?) and that I got used to handling that amount of money.
- (d) Once I was used to that, I went to \$2–20 for 2 hours.
- (e) Next was \$5–50 for 2 hours.
- (f) This was followed by \$25–200 for 3 hours.
- (g) Finally I went to \$50–500 (full scale) for 20 hours.

The idea, which has been valuable ever after, was to limit my bet size to a level at which I was comfortable and at which I could tolerate a really bad outcome.

Within each range, I used a fractional Kelly system, betting a percentage of my bankroll proportional to my expectation in favorable situations, up to the top of my current betting range, and the minimum of my current range otherwise. At \$50–500 (full scale) I bet full Kelly, which generally turned out to be a percentage of my bankroll a little less than the percentage expectation in favorable situations.

Convertible bonds

My next illustration is the evolution over more than two decades of a model for convertible bonds. Simplistically, a convertible bond pays a coupon like a regular bond but also may be exchanged at the option of the holder for a specified number of shares of the “underlying” common stock. It began with joint work during 1965 and 1966 with economist Sheen Kassouf on developing models for common stock purchase warrants. Using these models, we then treated convertible bonds as having two parts, the first being an ordinary bond with all terms identical except for the conversion privilege. We called the implied market value of this ordinary bond the “investment value of the convertible”. Then the value of the conversion privilege was represented by the theoretical value of the attached “latent” warrants, whose exercise price was the expected investment value of the bond at the (future) time of conversion.

Just as with warrants, we standardized the convertible bond diagrams so that the prices of different bonds versus their stock could be compared cross-sectionally on a single diagram. This was a crude first pass at finding underpriced issues. We also plotted the price history of

individual bonds versus their stock on the standardized diagrams. This was a crude first pass at finding an issue that was cheap relative to its history. You can see diagrams like these in *Beat the Market*. The scatter diagrams for individual convertibles were also useful in choosing delta-neutral hedge ratios of a stock versus its convertible bond.

I'd guessed the Black–Scholes formula in late 1967 and used it to trade warrants from late 1967 on. My progress in valuing convertibles quickened after I began to value the latent warrants using the model. Volatility, stock price and the appropriate riskless rate of return were now incorporated. But there remained the problem of estimating the future volatility for warrants (or options). There also was the problem of determining investment value. The future expected investment value of the convertible could only be estimated. Worse, it was not constant but also depended on the stock price. As the stock price fell, the credit rating of the bond tended to fall, reducing the bond's investment value, and this fact needed to be incorporated. We started with *ad hoc* corrections and then developed a more complete analytic model. By the time we completed this analytic work in the early 1980s, we had real-time price feeds to our traders' screens, along with real-time updated plots, calculations of alpha, hedging ratios, error bounds, and so forth. When our traders got an offer they could in most cases respond at once, giving us an edge in being shown merchandise.

To exploit these ideas and others, Jay Regan and I started the first market-neutral derivatives hedge fund, Princeton Newport Partners, in 1969. Convertible hedging was a core profit center for us. I had estimated early on that convertible hedging was a \$100 million idea for us. In the event, Princeton Newport Partners made some \$250 million for its partners, with half or more of this from convertibles. Convertible hedgers collectively have made tens of billions of dollars over the last three and a half decades.

Along the way we met a lot of interesting people: Names from the movie industry like Robert Evans, Paul Newman, and George C. Scott, Wall Streeters like Robert Rubin, Mike and Lowell Milken, and Warren Buffett, and academics like Nobelists Bill Sharpe, Myron Scholes, and Clive Granger.

Here are some of the things we learned about building successful quantitative models in finance. Unlike blackjack and gambling games, you only have one history from which to use data (call this the Heraclitus principle: you can never invest in the same market twice). This leads to estimates rather than precise conclusions. Like gambling games, the magnitude of your bets should increase with expectation and decrease with risk. Further, one needs reserves to protect against extreme moves. For the long-term compounder, the Kelly criterion handles the problem of allocating capital to favorable situations. It shows that consistent overbetting eventually leads to ruin. Such overbetting may have contributed to the misfortunes of Victor Niederhoffer and of LTCM.

Our notions of risk management expanded from individual warrant and convertible hedges to, by 1973, our entire portfolio. There were two principal aspects: local risk versus global risk (or micro versus macro; or diffusion versus jump). Local risk dealt with "normal" fluctuations in prices, whereas global risk meant sudden large or even catastrophic jumps in prices. To manage local risk, in 1973–1974 we studied the terms in the power series expansion of the Black–Scholes option formula, such as delta, gamma (which we called curvature) and others that would be named for us later by the financial community, such as theta, vega and rho. We also incorporated information about the yield "surface", a plot of yield versus maturity and credit rating. We used this to hedge our risk from fluctuations in yield versus duration and credit rating.

Controlling global risk is a quite different problem. We asked how the value of our portfolio would change given changes of specified percentages in variables like the market index, various shifts in the yield surface, and volatility levels. In particular we asked extreme questions: what if a terrorist explodes a nuclear bomb in New York harbor? Our prime broker, Goldman Sachs, assured us that duplicates of our records were safe in Iron Mountain. What if a gigantic earthquake hit California or Japan? What if T-bills went from 7% to 15%? (they hit 14% a couple of years later, in 1981). What if the market dropped 25% in a day, twice the worst day ever? (it dropped 23% in a day 10 years later, in October 1987. We broke even on the day and were up slightly for the month). Our rule was to limit global risk to acceptable levels, while managing local risk so as to remain close to market neutral.

Two fallacies of which we were well aware were that previous historical limits on financial variables should not be expected to necessarily hold in the future, and that the mathematically convenient lognormal model for stock prices substantially underestimates the probabilities of extreme moves (for this last, see my columns in *Wilmott*, March and May 2003). Both fallacies reportedly contributed to the downfall of LTCM.

Two questions about risk which I try to answer when considering or reviewing any investment are: “What are the factor exposures”, and “What are the risks from extreme events?”.

Statistical arbitrage

Hedging with derivatives involves analytical modeling and, typically, positions whose securities will have known relationships at a future date. A rather different modeling approach is involved with a product called “statistical arbitrage”.

The key fact is the discovery of an empirical tendency for common stocks to have short-term price reversal. This was discovered in December 1979 or January 1980 in our shop as part of a newly initiated search for “indicators”, technical or fundamental variables which seemed to affect the returns on common stocks. Sorting stocks from “most up” to “most down” by short-term returns into deciles led to 20% annualized returns before commissions and market impact costs on a portfolio that went long the “most down” stocks and short the “most up” stocks. The annualized standard deviation was about 20% as well. We had found another potential investment product. But we postponed implementing it in favor of expanding our derivatives hedging.

For the record, we had also been looking at a related idea, now called “pairs trading”. The idea was to find a pair of “related” stocks that show a statistical and perhaps casually induced relationship, typically a strong positive correlation, expecting deviations from the historical relationship to be corrected.

Meanwhile at Morgan Stanley, the brilliant idiosyncratic Gerry Bamberger (the “unknown creator” of MS mythology) discovered an improved version in 1982: hedge within industry groups according to a special algorithm. This ran profitably from 1982 or 1983 on. However, feeling underappreciated, Bamberger left Morgan in 1985, co-ventured with us, and retired rich in 1987. As a result of his retirement, we happened to suspend our statistical arbitrage operation just before the 1987 crash. We restarted a few months later with our newly developed factor-neutral global version of the model.

Too bad: simulations showed that the crash and the few months thereafter were by far the best period ever for statistical arbitrage. It was so good that in future tests and simulations we had to delete this period as an unrealistically favorable outlier.

Statistical arbitrage ran successfully until the termination of Princeton Newport Partners at the end of 1988.

In August 1992 we launched a new, simpler principal components version. This evolved into the “omnivore” program, which incorporated additional predictors as they were discovered. The results: in 10 years, from August 1992 through October 2002, we compounded at 26% per annum net before our performance fee, 20% net to investors, and made a total of about \$350 million in profit. Some statistics: 10 day average turnover; typically about 200 long and 200 short positions; 10,000 separate bets per year, 100,000 separate bets in 10 years. The gross expectation per bet at about $(2/3)\% \times 1.5 \text{ leverage} \times 2 \text{ sides} \times 25 \text{ turnovers per year}$ is about 50% per year. Commissions and market impact costs reduced this to about 26%.

Concluding remarks

Where do the ideas come from? Mine come from sitting and thinking, academic journals, general and financial reading, networking, and discussions with other people.

In each of our three examples, the market was inefficient, and the inefficiency or mispricing tended to diminish somewhat, but gradually over many years. Competition tends to drive down returns, so continuous research and development is advisable. In the words of Leroy Satchel Paige, “Don’t look back. Something might be gaining on you”.