Basic Principles of Computerized Investment Analysis

By

Dr. Robert C. Smithson

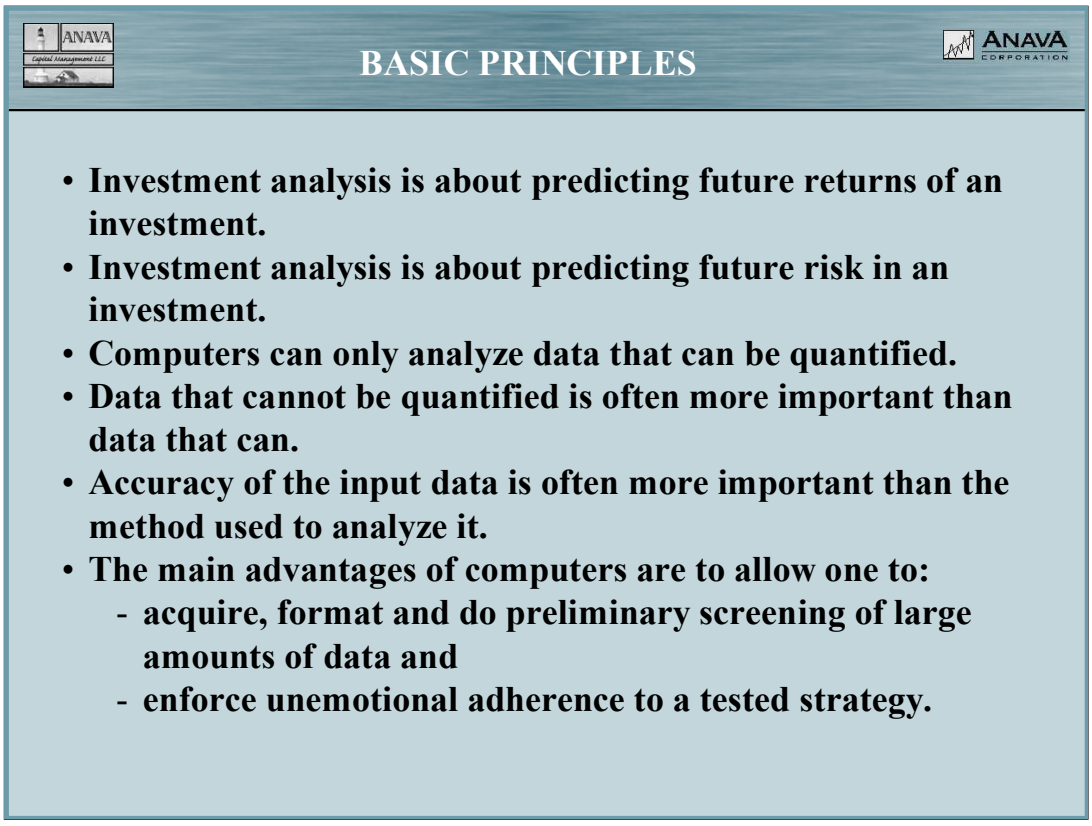
Anava Capital Management LLC
Cupertino, CA
408-918-9333

Computerized investing takes advantage of the power of modern computers to enable investors to make good investment decisions. However, the results of computerized investing are only as good as the programs that are used and the data that they analyze. Computers are useful for the following tasks:

1. keeping track of holdings, expenses, profits and losses, and other similar bookkeeping tasks
2. finding and storing information related to current and potential investments
3. predicting the future performance of your current and potential individual investments
4. portfolio optimization based on the results of expected performance of individual investments
5. making specific buy and sell decisions based on the above

Of these, task 3, predicting the future performance of an investment, is the most controversial and difficult (some would say impossible). This series of two classes will cover methods by which this problem can be approached with the aid of computers. We will also look at the results that have been achieved with these methods and to what degree they can be relied upon.

The extent to which individual investors are willing and able to take advantage of computers for these tasks depends not only on the degree of computer literacy of the individual involved, but more importantly on the personality of that investor. Therefore the necessary attitudes to adopt in order to use computerized analysis will also be discussed.



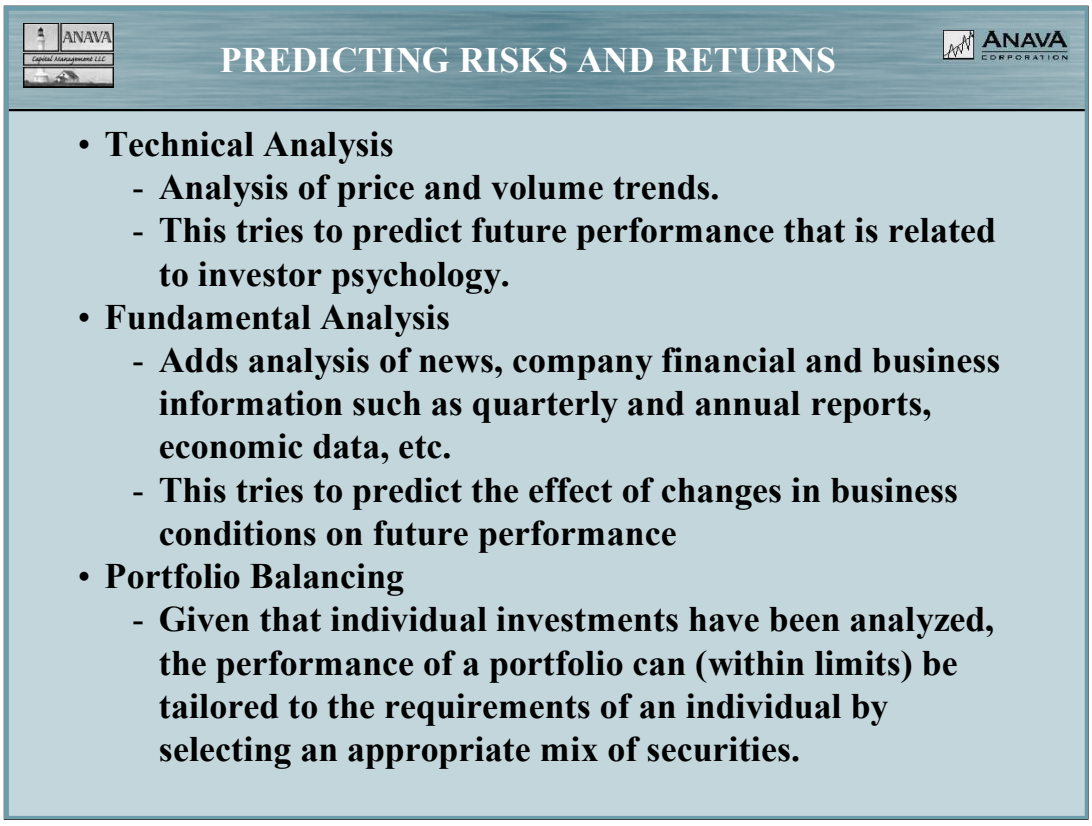
The objective of all investment analysis, computerized or not, is to predict the future returns and risks of a particular investment. In this case, an investment is defined as something that can be purchased or sold which is capable of generating positive net returns for the owner. This is a rather broad definition of the word "investment" that includes both categories of incoming-producing things: those which can produce income merely by owning them (e.g. – CD's, stocks) and those which require additional labor or cash to produce a profit (e. g. – a milling machine, or a house with net negative cash flow).

Predicting the future is always risky, and a prediction of the future without at least some idea of the probability that the prediction is correct is absolutely useless for investment purposes. In fact, in order to be useful, the probability must be explicitly stated in a quantitative way. This is one place where computers can help.

Unfortunately for computerized analysis, much of the most important data that determine the future performance of investments are not easy to quantify. Also, it can be difficult to obtain accurate data, even when available. One of the problems with computerized screens is that sometimes data entry errors are made that can make the screen results invalid. A greater problem is that the data that is available necessarily relates to an investment's performance in the past, not in the future – thus the universal warning to investors that "past performance does not necessarily reflect future results." Luckily, this is true for everyone in the market, so at least the playing field is level.

Computerized investing has two main advantages:

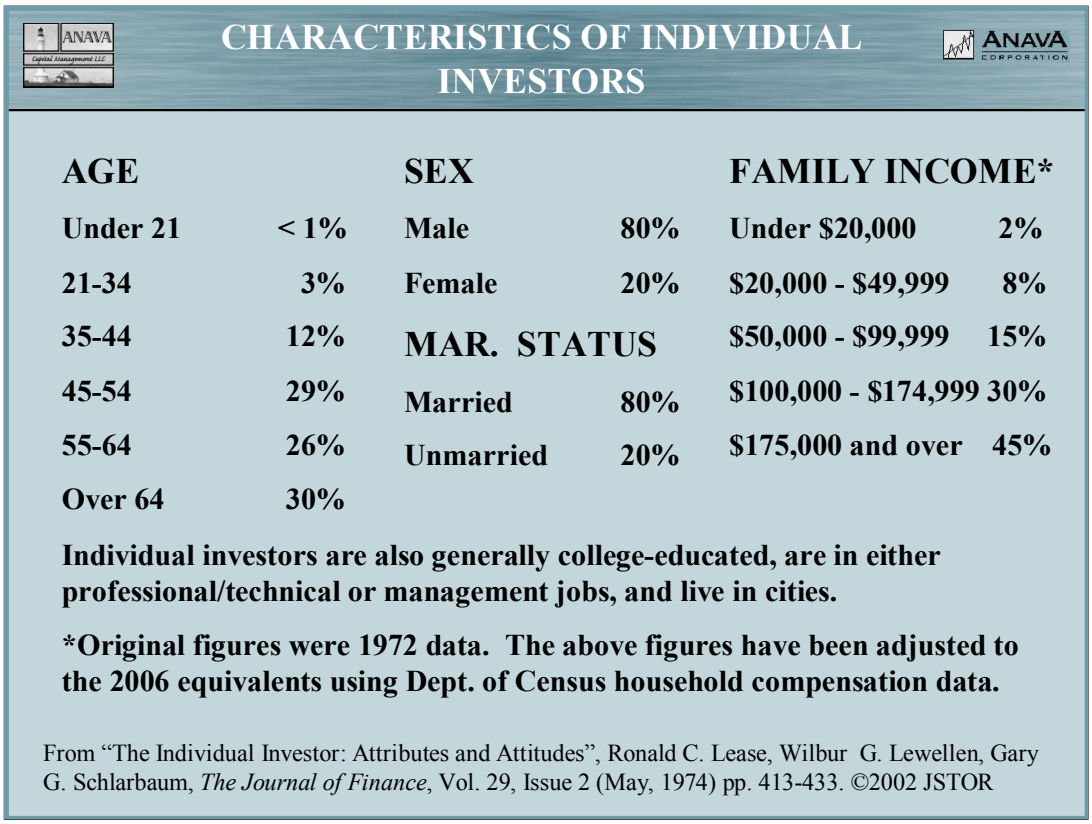
First, computers can screen thousands of stocks in seconds, allowing individuals to find the investments that best meet their objectives without depending on their broker's research staff. Because even large brokerage houses concentrate on a relatively small number of investments which are approved for recommendation to clients, this potentially allows individuals to do a better job than the broker in building a portfolio designed for their needs. Second, because computers are quantitative devices, they force investors to quantify both their investment goals and their assessments of specific investments. This is important because it can dramatically reduce the effects of the two most important reasons for market losses – fear and greed.



The value and returns of investments change for two basic reasons. Short term changes in value are mostly caused by changes in investor perception of the desirability of an investment. Long-term changes are mostly caused by real changes in the earning power of the business represented by the investment. An examples of the former is the bandwagon effect as a result of panic or greed (Stocks are risky! Get out of stocks!). An example of the latter is a specific business development affecting the probable future of a company (Recent headline: "Graham Corporation Awarded $3.7 Million U.S. Oil Refinery Order").

Some investors prefer to use technical analysis, which tries to predict investor psychology by watching price and volume fluctuations and other indicators of past investor sentiment. They belief that certain patterns in past investor behavior are useful in predicting future behavior and thus price changes caused by investor sentiment changes. Others prefer fundamental analysis which looks at financial figures such as sales, earnings and other concrete measures of business health to predict price changes based on actual changes in the business fortunes of the company. Interestingly, until recently, most investors concentrated on one or the other of these methods, and mostly ignored the other. That has changed with the advent of computerized quant investing. "Quant" is short for "quantitative."

Portfolio balancing is not the subject of this talk, but is mentioned for completeness. The optimum method of investing rarely if ever involves making just one investment. That is mostly because unforeseen things can happen to any one investment. Thus having at least ten with equal dollars in each or so means that if something goes badly wrong with one, the loss is limited to ten percent. Portfolios are for downside protection.



The above data shows that individual investors are mostly old, affluent, male and married. The numbers shown are quite close to those we have seen in responses to questionnaires returned by attendees at the 2006 AAII seminar on Computerized Investing. It that respect at least nothing much seems to have changed.

RETURNS FROM INDIVIDUALS' STOCK INVESTMENTS BEFORE COMPUTERS

Data for the common stock transactions of 2,506 individual investor accounts from 1964 through 1970:

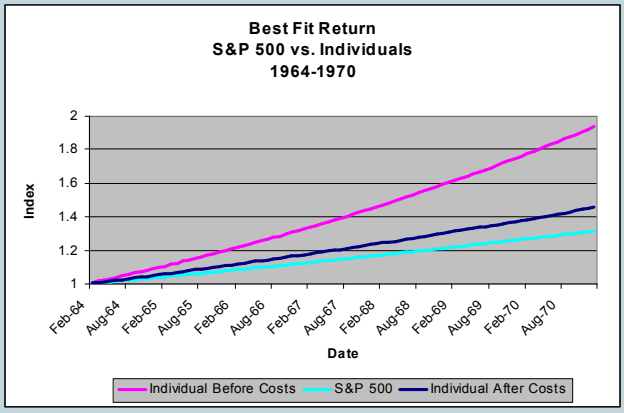
Short Positions: 3% **Long Positions 97%**

Individual Mean Annual Return (%) Before / After Commissions:

9.9 / 5.5

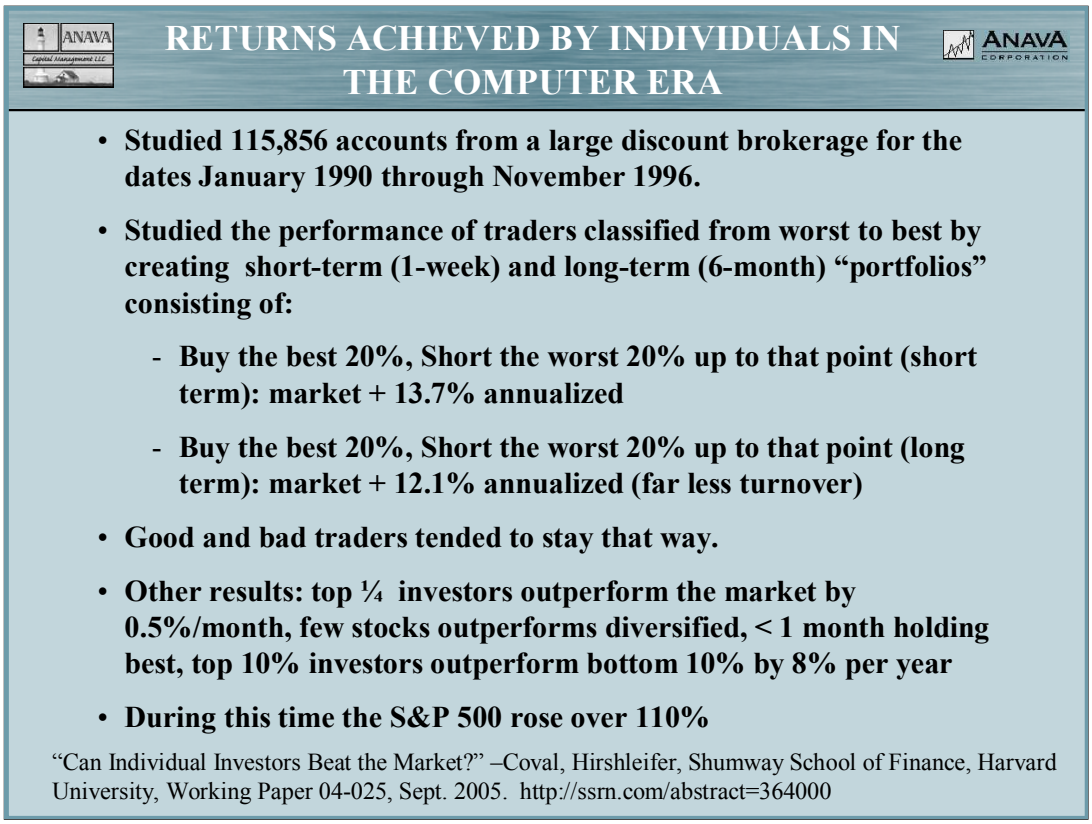
S&P 500:

3.6



"Realized Returns on Common Stock Investments: The Experience of Individual Investors," Gary G. Schlarbaum, Wilbur G. Lewellen, & Ronald C. Lease, *The Journal of Business,* 1978, vol. 51. No. 2

The above figures were compiled for the years 1964 through 1970, years in which the use of computers by individuals was essentially non-existent. The Internet did not exist at that time, although its predecessor, ARPANET, which was developed for military use, did exist in a rudimentary form by 1970. The first personal computers appeared in the mid 1970s. At the time the conventional wisdom was that individuals "always lost money." It seems not to have been true.

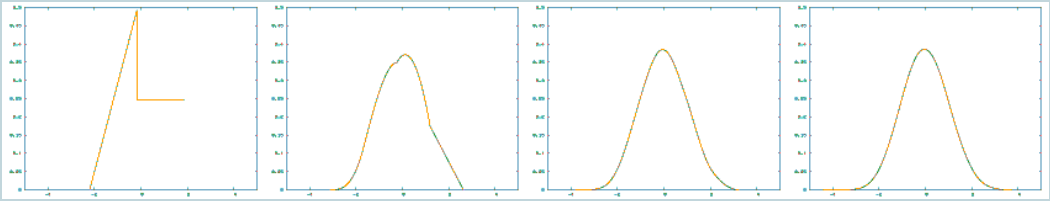


This is a study that showed that the best individual investors out performed the market substantially, even if individuals as a group did poorly. The question here is how to put yourself in the top 20%. The authors suggest that brokerages with access to individual trading records can do it by duplicating the study. That is, they should recommend the same stocks that their most successful accounts are buying, and short the stocks that the least successful accounts are buying. That's not a strategy an individual should use, and it's likely illegal for a brokerage. Still, it does indicate that some people know what they are doing and others don't, and that those who do make more money. For this study it was particularly true for the short term, but held true for longer term trades as well.

Keep in mind that times and market conditions change, and what was true for the boom times of the 90s may not be true any more.

These guys lost money on a lot of weeks!

WHAT'S WRONG WITH CONVENTIONAL STATISTICS

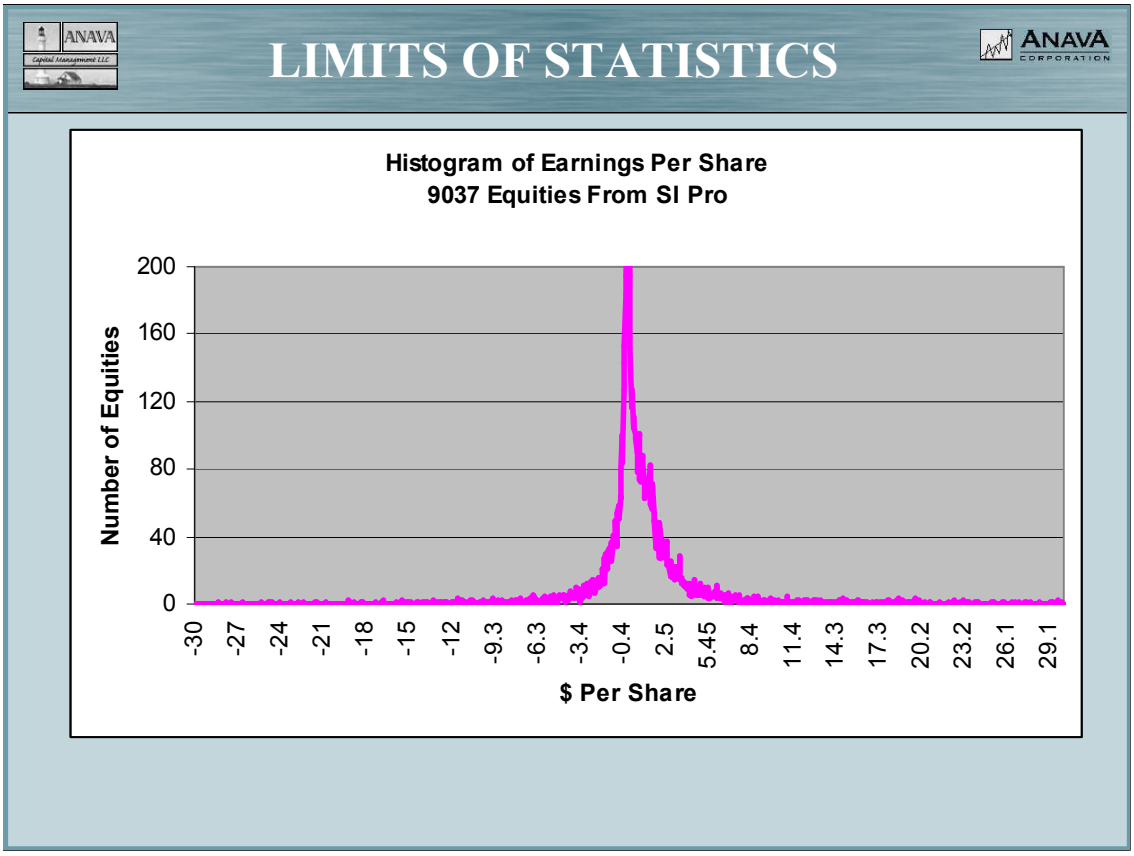


1 SAMPLE | 2 SAMPLES | 3 SAMPLES | 4 SAMPLES

- **THE CENTRAL LIMIT THEOREM**
  - **Let X_1, X_2, X_3, ... be a set of *n* independent and identically distributed *random* variables having finite values of mean μ and variance σ^2. The central limit theorem states that as the sample size *n* increases, the distribution of the sample average approaches the normal distribution with a mean μ and variance σ^2/n *irrespective of the shape of the original distribution.***
- **NORMAL DISTRIBUTION (PROBABILITY DENSITY FUNCTION)**
  - $$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

http://en.wikipedia.org/wiki/Central_limit_theorem

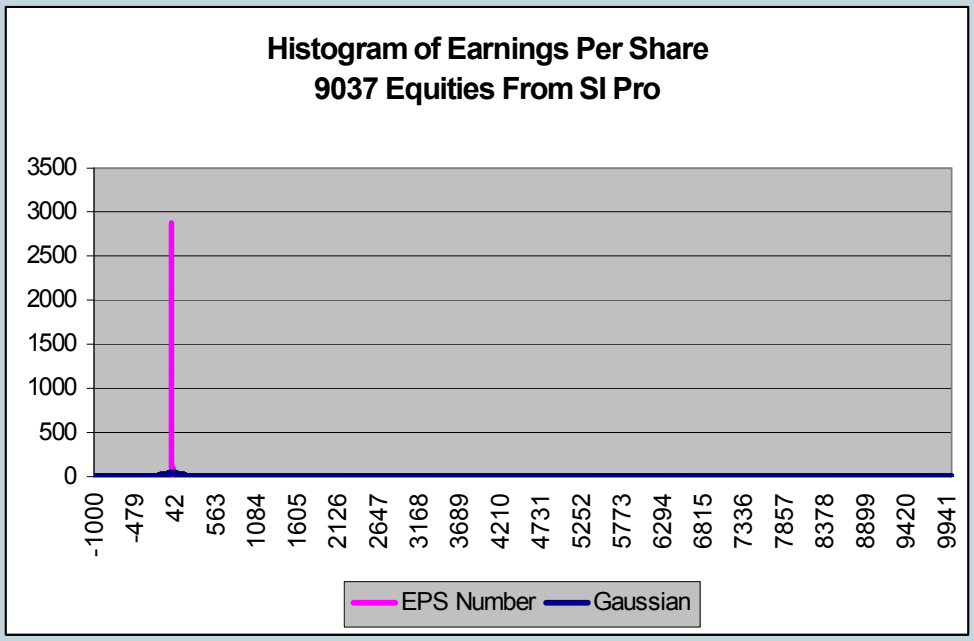
Conventional statistical error analysis assumes the central limit theorem. It assumes the normal distribution and talks about things like two-sigma and three-sigma events. Then it tells you that there's about a 98% probability that the value of the number being analyzed will lie within the three-sigma limit based on that assumption. This can lead to serious errors in investing.



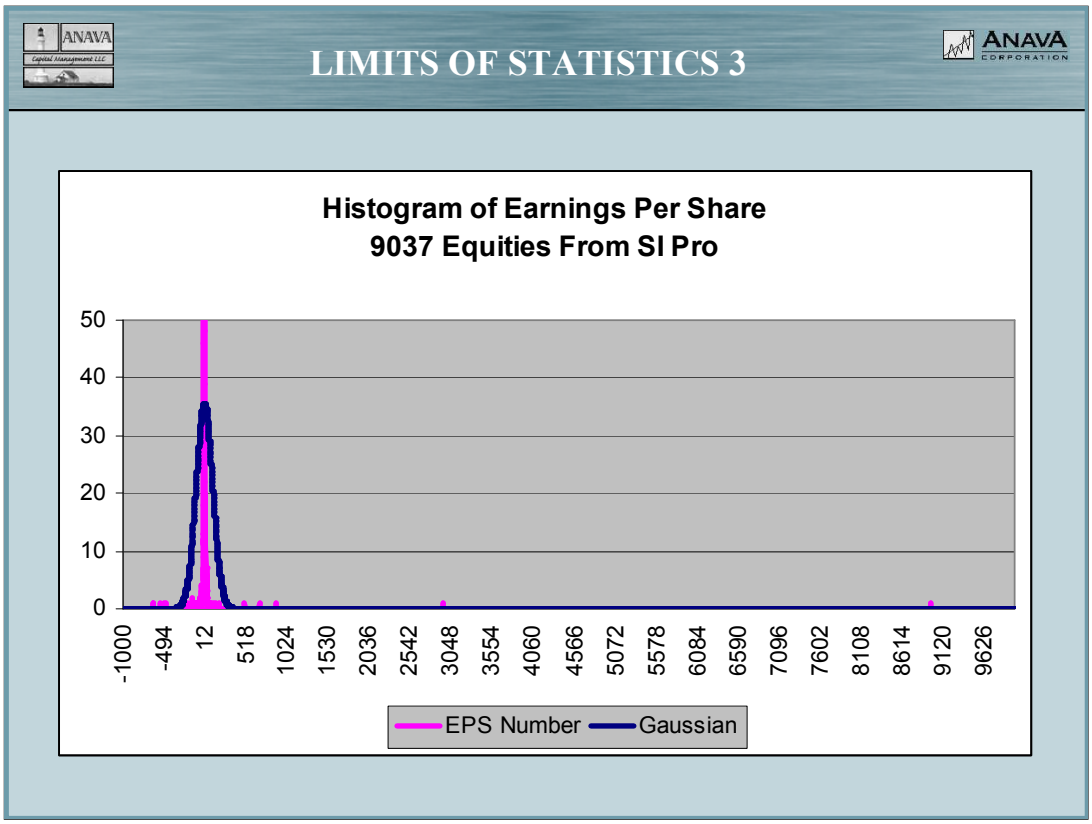
This is a portion of a histogram that shows earnings per share as the number of companies (out of the 9037 company SI Pro database) that have a given EPS value. It looks rather like a normal distribution, but has a disturbingly large number of points in the tails, it's got a much sharper peak than a normal distribution and is asymmetrical. There are more companies with positive earnings than negative. (Guess the world won't end just yet.)

Why are these things true? Because people don't act at random. Most all of the people responsible for these investments are trying to make money, hence the asymmetry.

LIMITS OF STATISTICS 2



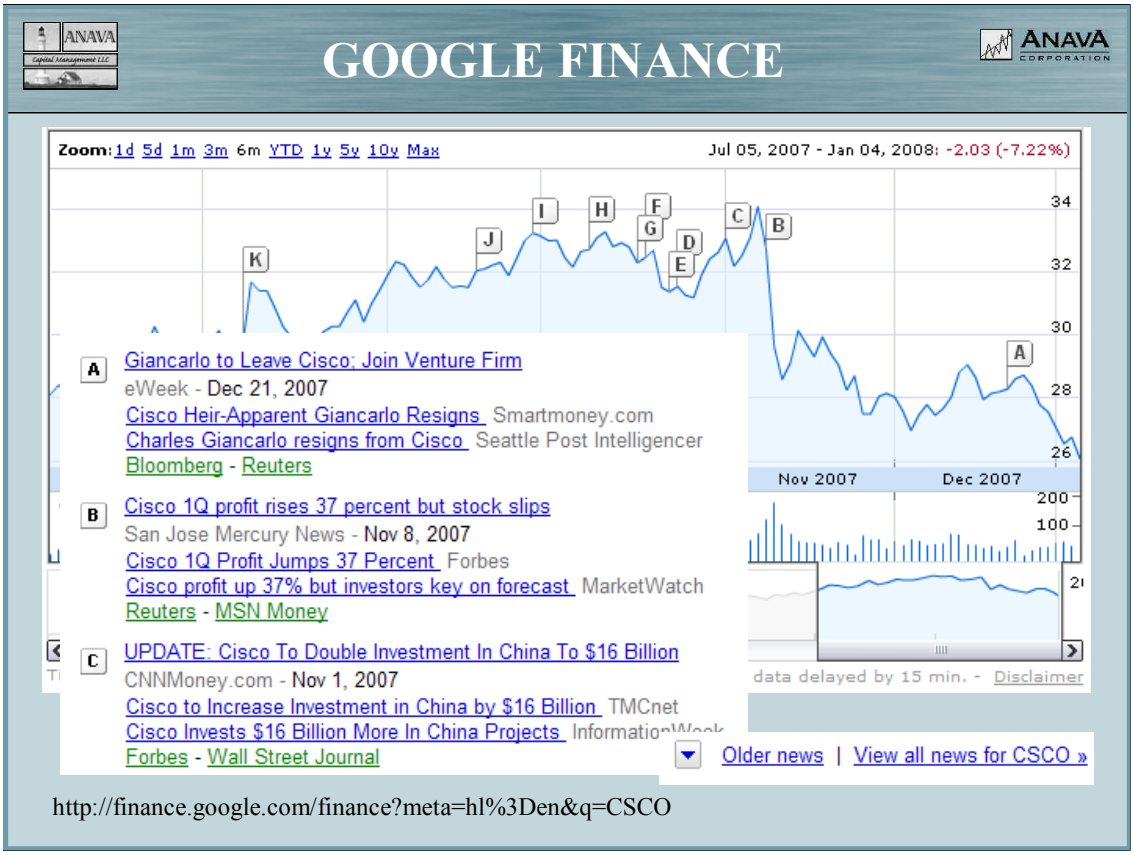
Here's the whole histogram shown in the previous slide, with a normal distribution with the same mean and standard deviation superimposed. It is NOT a good fit. Why? Look at the next slide.



See those single example outliers? They have produced a standard deviation that produces a normal fit that is much too wide for the actual peak. The normal distribution has the same mean, standard deviation, and integral as the histogram, but it is low and wide near the peak, and drops to near zero far before the outliers disappear.

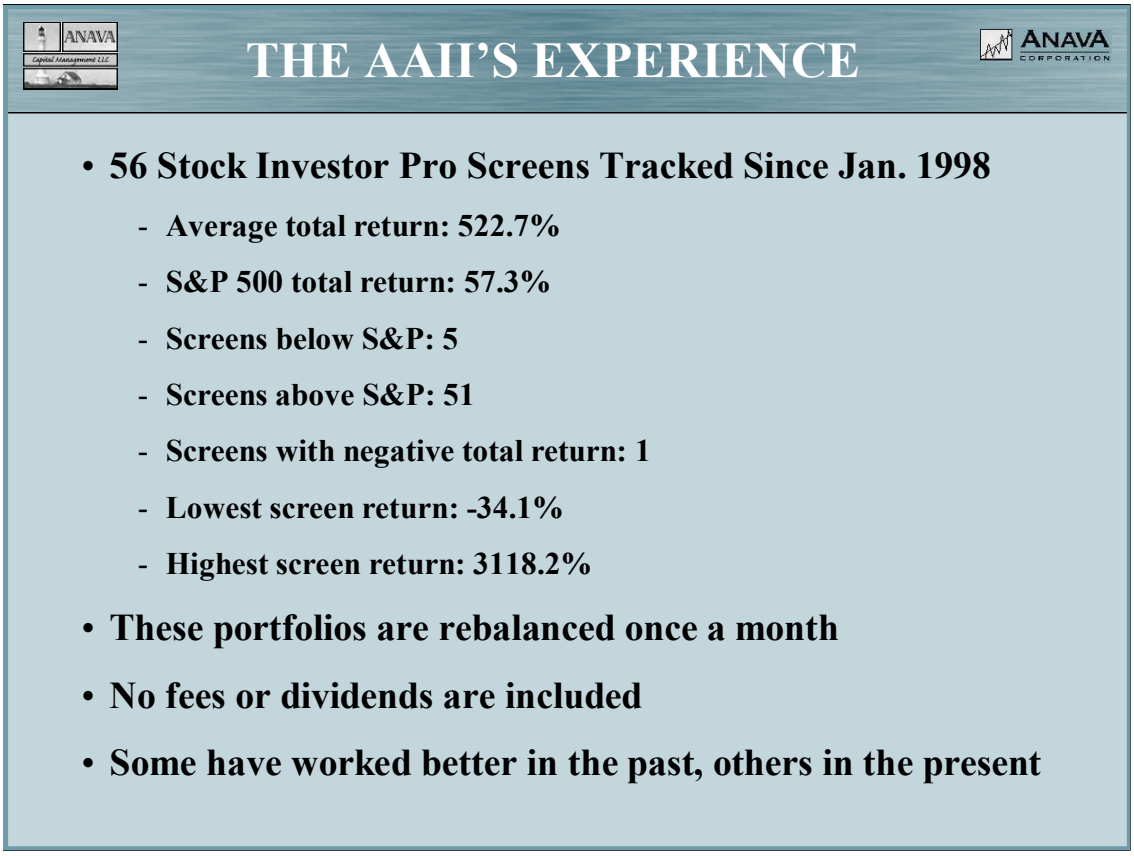
HOW TO FIX THE PROBLEM

- **Conventional Statistics work reasonably well *if* you:**
  - **Throw out the outliers**
  - **Pick your investments so you have as few surprises as possible**
  - **Diversify to protect yourself from surprises that happen anyway**
  - **Come to the next class where we'll talk about how to do those things.**
- **Another approach is "data-driven" statistics:**
  - **There are no assumptions about the nature of the probabilities**
  - **Instead of a normal distribution empirical histograms are used directly**
  - **This approach works well *if* you have enough data**
  - **We'll talk about that next time too.**



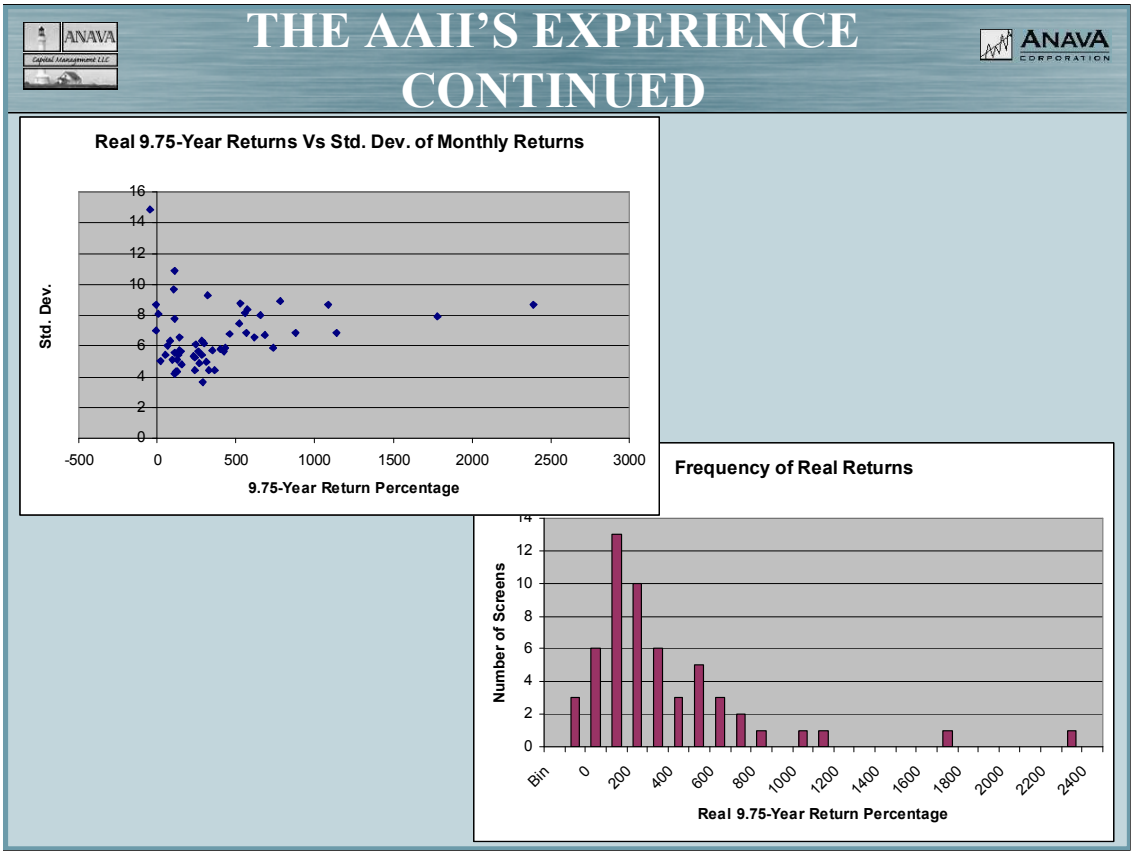
http://finance.google.com/finance?meta=hl%3Den&q=CSCO

This is an example of how the use of a computer and the Internet can allow a type of analysis that has proven very profitable in the past. The old way was to clip articles out of newspapers and tape them onto a gain chart of a stock's price attached to a wall. Then read the articles and see if there is a correlation between a certain type of article and stock price changes. Google Finance has automated this process as a part of the current beta version of their charting site. The chart is at http://finance.google.com/finance?meta=hl%3Den&q=CSCO



The chapter has an Active Stock Investor Pro users' group run by Andy Prophet. See our chapter website for more information and meeting schedules. The website can be found at www.siliconvalleyaaii.org.

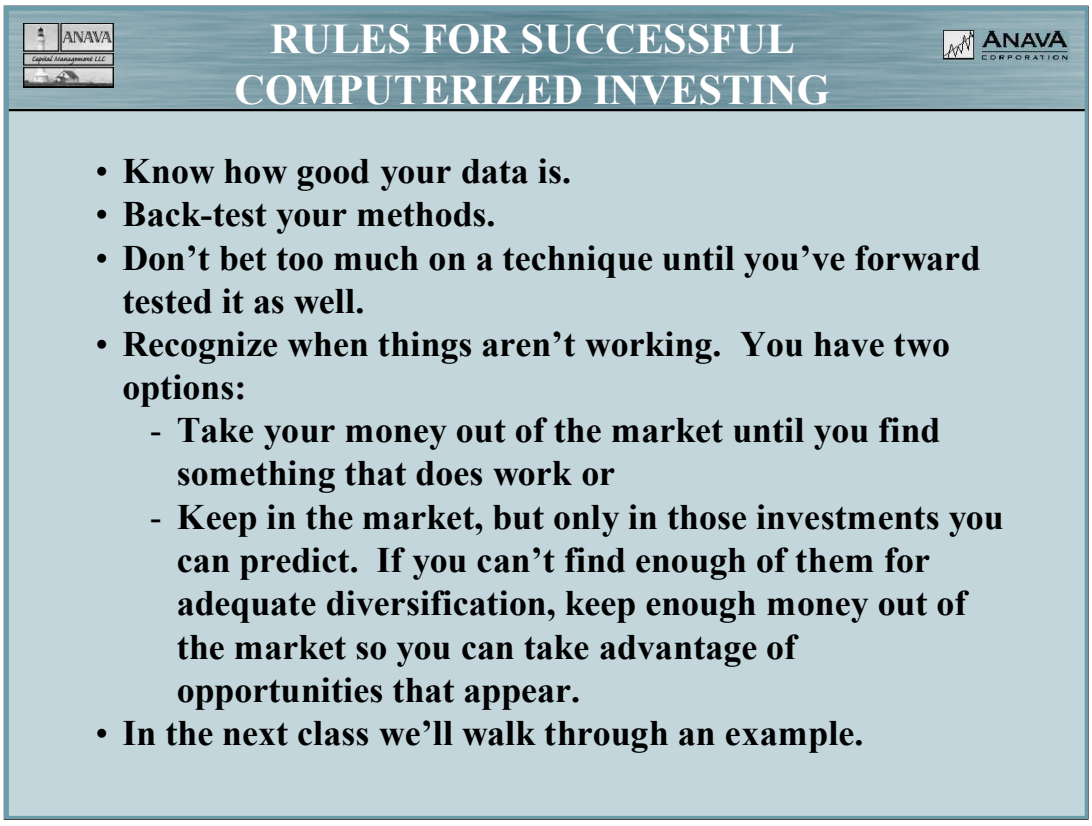
These screening results are from the AAII website at http://www.aaii.com/stockscreens/performance.cfm.



The scatter plot shows a general confirmation of the conventional wisdom that higher returns go with higher volatility. However the effect is certainly not linear. Most returns near zero show a monthly standard deviation between 4 and 6, while a return of almost 2500% is at about 8.5. In addition, there are a significant number of low-return screens that show very high standard deviations, including the only screen with a significant loss at about 15.

The histogram shows the screens clustering around real returns of about 200% over the 9.75-year period. Anything above 1000% is probably a "fat tail" outlier. However, are these high performers just chance? The highest is a new screen, the O'Shaughnessy Tiny Titans, and the second highest is the Zweig screen. The newer screens were back-tested in a way that should eliminate most survivorship bias. (If you test screens on earlier data on presently existing companies only, you eliminate those that no longer exist from consideration. However, the AAII has old data available in SI Pro, and therefore can, and presumably did, test on the actual data that was available at the earlier date) Nevertheless, recently developed screens were done by people who know what worked in the past, and tend to give unrealistically good results when back-tested. The Tiny Titans is a recently added screen, so the performance is suspect.

Many of the screens in this data date back to, or near to, 1998, the earliest year used here. They have mostly been tested in real time, after the screen was developed. The second-best screen, the Zweig screen, is one of these. It has never had a negative year. Is this just chance or is there something valid here?



In the next class we'll work through an example with some spreadsheets that allow you to do what we have been talking about.