Cramer vs. Pseudo-Cramer

Patrick Burns*

3rd October 2007

Abstract

A recent *Barron's* article examined the efficacy of stock recommendations on the television show *Mad Money*. Statistical analyses of stock recommendations are scrutinized here in detail, and a powerful analysis using random portfolios is suggested. Differences between simple returns and log returns are discussed, as is the usefulness of the statistical bootstrap. The cost to individuals of trading stocks can easily overwhelm even quite good recommendations.

1 Introduction

An article entitled "The Cramer Effect—and Defect" by Bill Alpert appeared in the August 20, 2007 edition of *Barron's*. The article (which is available on http://www.barrons.com) explored the stock-picking ability of Jim Cramer on the CNBC show *Mad Money*. The article concluded that his market-beating ability is questionable—a conclusion that predictably was at odds with other opinions, including those of CNBC.

The "Cramer effect" refers to the phenomenon that stocks recommended on *Mad Money* experience a large return the day after the broadcast. The average jump is on the order of 2 percent.

I was the advisor on the analysis of the data for the *Barron's* article. This paper discusses the strengths and weaknesses of a number of analyses that have been performed. It finishes with an analysis—proposed but not implemented—that is substantially better than any that were performed, and that gives the paper its title.

2 A Word about Returns

There are two main types of return: simple returns and log returns. Another name for log returns is "constantly compounded returns". There are numerous

*This paper is available in the working papers section of http://www.burns-stat.com. The author thanks Bill Alpert and Oliver Graham for useful comments.

alternative names for simple returns, including "real return". If there is no indication at all of what type of return it is, it is often a simple return.

If P_1 is the price of the asset at time 1 and P_2 is the price at time 2, then the simple return from time 1 to time 2 is:

$$R_2 = \frac{P_2 - P_1}{P_1} = \frac{P_2}{P_1} - 1 \tag{1}$$

The formula for a log return is:

$$r_2 = \log\left(\frac{P_2}{P_1}\right) = \log(P_2) - \log(P_1) \tag{2}$$

where "log" means the natural logarithm.

Note the reasonably common convention that uppercase R means a simple return, and lowercase r means a log return.

Returns are unit-less—the currency in the denominator cancels the currency in the numerator. However, they do pertain to a period of time. Returns are often annualized, you want to know if they have been. You also need to know if the returns are expressed in percent or not (whether simple or log returns).

Simple returns can be arbitrarily large, but can not be less than minus one. A simple return of -1 means that all of the money has been lost. Log returns can take on any number. As Figure 1 shows, log returns are always less than simple returns except they are equal at zero. At the extreme a simple return of -1 corresponds to a log return of negative infinity.

However, for short periods of time (such as a week or less), the difference will virtually always be very small. A log return of 1% corresponds to a simple return of 1.005%. A log return of 10% is equivalent to a 10.52% simple return.

If you have one type of return, you can easily get to the other type. The formula to convert from a simple return to a log return is:

$$r_2 = \log(R_2 + 1) \tag{3}$$

To go from log returns to simple returns, do:

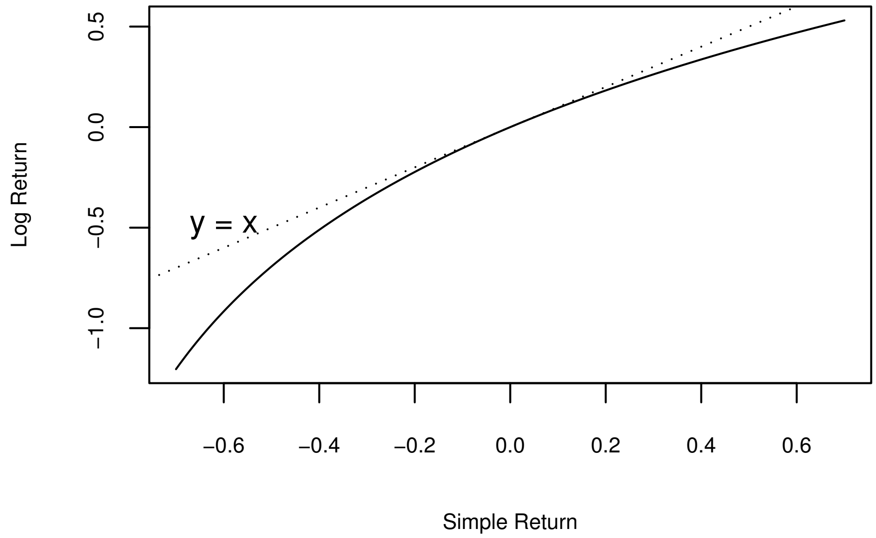
$$R_2 = \exp(r_2) - 1 \tag{4}$$

These formulas can be very useful when doing operations on returns. You can transform to the type of return that is easiest for the operation and then transform back at the end.

Simple returns are easy when going from individual assets to a portfolio. The simple return of the portfolio is the weighted sum of the assets' simple returns.

The log return of a long period of time is the sum of the log returns of periods that partition the period. Consider having three time points (and hence two periods). The log return for the whole period is:

Figure 1: The value of simple returns relative to log returns, with the y=x line as a reference.



$$\log\left(\frac{P_3}{P_1}\right) = \log\left(\frac{P_2 P_3}{P_1 P_2}\right) = \log\left(\frac{P_2}{P_1}\right) + \log\left(\frac{P_3}{P_2}\right) \quad (5)$$

This technique can obviously be extended to any number of intermediate time points.

Summing over time is a reason to believe that log returns follow a normal distribution. The reasoning uses the Central Limit Theorem about the sum of a large number of similar random variables. Though the assumption of a normal distribution is often made, it is decidedly not true. Returns have a much higher probability of extreme values than the normal distribution.

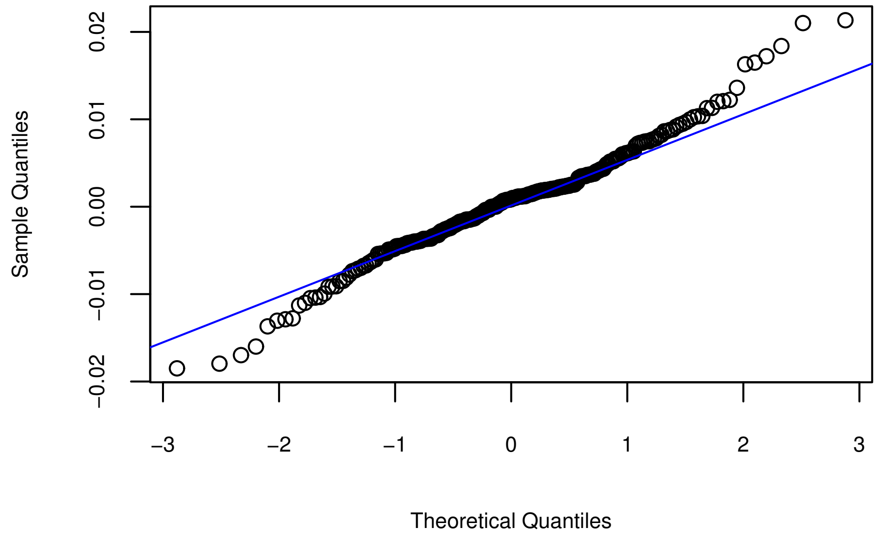
Figure 2 shows a normal qqplot of the 2006 daily returns of the S&P 500 index. If the returns were normally distributed, they would lie close to the line. You can see how improbable the assumption of normality is by using the command:

```
rn <- rnorm(251); qqnorm(rn); qqline(rn, col='blue')
```

a number of times in the R language [R Development Core Team, 2007]. The random normals tend to cling to the line more than the 2006 S&P returns. The S&P returns for the exceptionally non-volatile years 2004 and 2005 do look normal. The 2007 returns look fatter tailed (less normal) than 2006.

To summarize the uses of the two types of returns, simple returns often make more investment sense while log returns often are better for statistical modeling.

Figure 2: Daily returns of the S&P 500 during 2006 compared to a normal distribution.



3 Data

A few datasets have been used in analyses of Cramer's record:

- 30 picks selected in [Nayda, 2006].
- 1275 buy recommendations over the history of *Mad Money* compiled by YourMoneyWatch.com.
- A six-month history of all recommendations (3458 in number) maintained by TheStreet.com. Jim Cramer founded TheStreet.com but claims these data are unofficial.
- A dataset of 445 buy recommendations in 2007 selected by CNBC.

4 Some Analyses

Statistical textbooks often imply that there is one correct analysis. The reality is that all analyses are wrong. The real task is to find an analysis from among those that are least wrong.

The analyses for *Barron's* and the graphics for this paper were all done with R [R Development Core Team, 2007]. This is software that is purpose-built for data analysis. Many people think it is the best data analysis software that money can buy, except that R is free.

4.1 Descriptions and Conclusions

[Engelberg et al., 2007] studied 391 first-time *Mad Money* recommendations made between 16 November 2005 and 23 June 2006, along with a variety of other data to study a number of questions. While their analyses are very interesting, they are not especially similar to those presented here and so will not be discussed further.

Analysis N

The Nayda analysis [Nayda, 2006] selects the first company recommended during the Smart Money segment of the broadcast from 1 November 2005 to 12 December 2005. The stock returns were compared to the S&P 500 returns.

If a selection is to be made, this is a quite good one to make. People tend to remember the first (and the last) items in a presentation. So viewers are probably more likely to act on the first recommendation than any others. Also this is presumably the recommendation that Cramer thinks is his strongest for the day.

The result was that the recommendations—when bought the day after the broadcast and held for either 19 days or 39 days—underperformed the S&P 500.

Analysis B1

The first question *Barron's* asked was in relation to the Cramer effect. What is the stock's behavior just before and just after the mention of the stock on *Mad Money*? To answer this question the data were adjusted for the "market" (the S&P 500) and the stock's behavior relative to the market the previous year.

Figure 3 is a result from that study. It shows that on average the stocks rise mildly (relative to the market and past performance) before the broadcast, rise a lot the day after, and then slide downward for a number of days.

Analysis B2

The other main question was how viewers might fare when implementing *Mad Money* recommendations. For this the stock returns were adjusted for the market but not for prior behavior of the stock.

Figure 4 shows the mean cumulative log return similar to Figure 3. There are two differences between Figure 3 and Figure 4. The first is that the cumulative returns are from day 0 rather than day minus 2—this is of no consequence because the starting point doesn't change the behavior. (Day 0 is the day of the broadcast, but markets have closed before the broadcast is aired.) The second difference is that Figure 4 does not adjust for the stock's previous year. Note that adjusting by the previous year's behavior pulls the returns downward. In other words Cramer's recommendations tend to have positive momentum relative to the S&P 500.

Figure 3: The cumulative log return from two days before broadcast relative to both the S&P 500 and the stock's prior year. The pointwise 95% confidence interval is indicated in yellow. (YourMoneyWatch data)

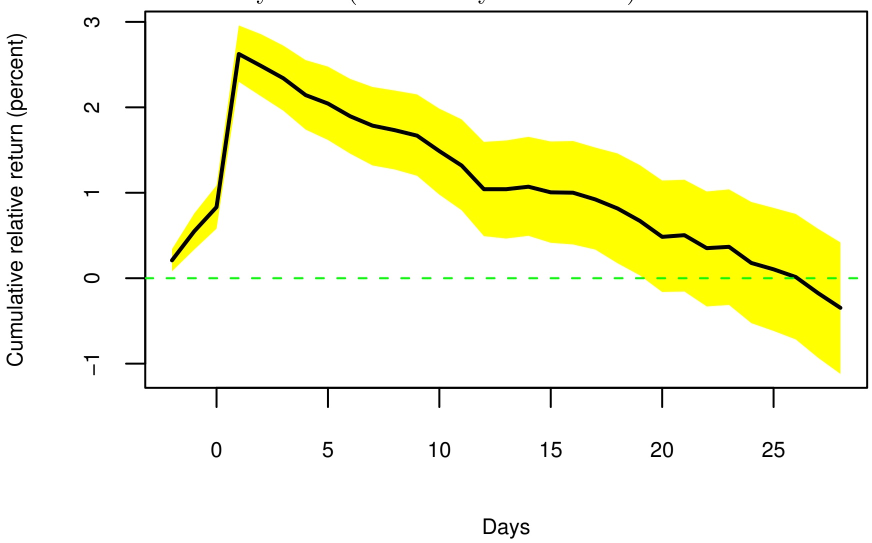


Figure 4: The cumulative log return from the day of the broadcast relative to the S&P 500. The pointwise 95% confidence interval is indicated in yellow. (YourMoneyWatch data)

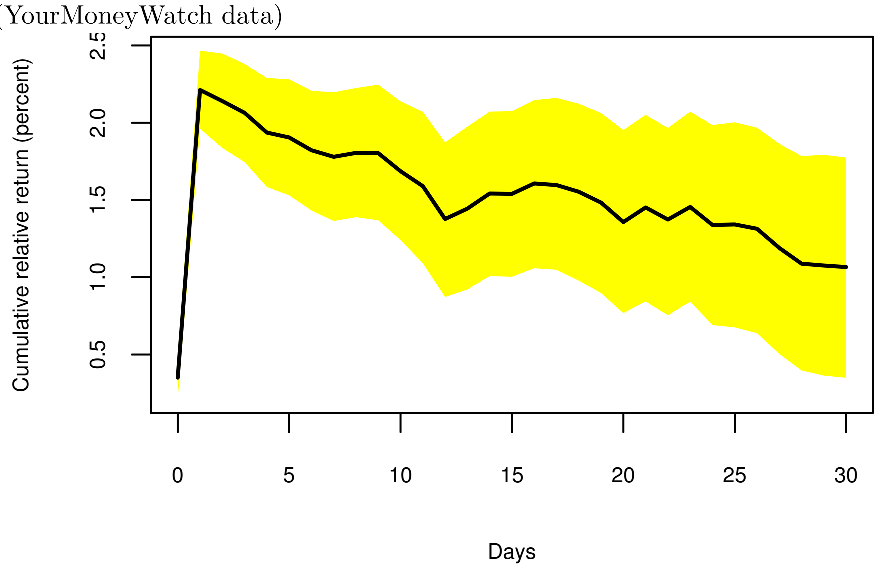


Table 1: 95% confidence interval of the mean simple return relative to the S&P 500 for Featured positive recommendations bought on day 5 and held 25 days. (TheStreet data)

n. obs.	mean	lower bound	upper bound
505	0.74%	-0.06%	1.60%

Analysis C

CNBC provided an analysis of a selection of recent, emphatic recommendations. They buy five days after the broadcast and hold for either one or two months. Given Figure 4 it is sensible to wait to buy rather than buying the day after the broadcast.

The mean return of the recommendations was 1.18% for one month and 2.50% for two months while the average one month return of the S&P 500 over the period was 0.39%. Given that large of a difference they declared it significant.

4.2 Key Considerations

4.2.1 Significance

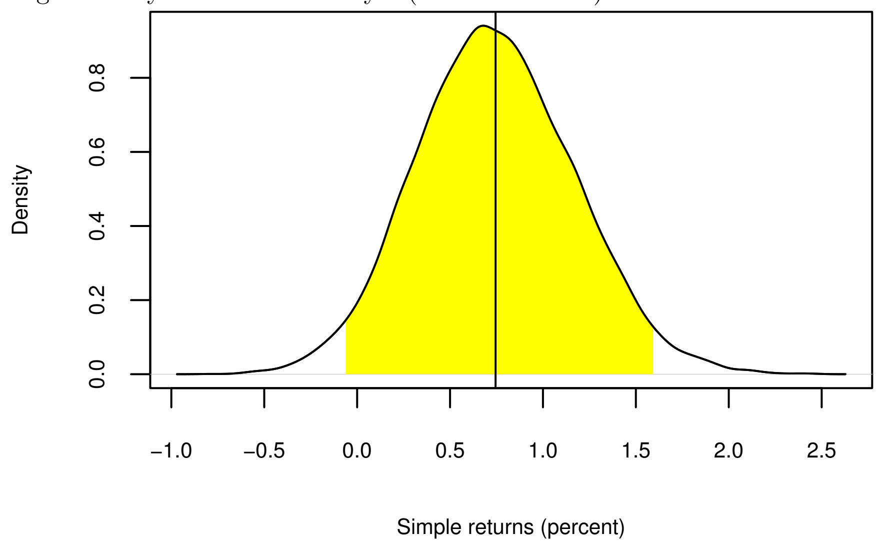
An analysis that just produces estimates is incomplete. The variability of the estimates also needs to be addressed. This is especially important in this context—stock returns are very variable, so what seems like large differences may have minimal statistical significance.

The Nayda analysis does not explicitly address significance. It is put into a statistical hypothesis framework in which the alternative hypothesis is that the *Mad Money* recommendations outperform. Since the sample results do not outperform, it is obvious that the null hypothesis will not be rejected at typical levels of significance, so significance is at least implicitly addressed. (All introductory statistics textbooks have descriptions of hypothesis tests.)

Confidence intervals are a form of significance that is often preferred to hypothesis tests. Confidence intervals indicate a range that the value might fall into. Table 1 shows a 95% confidence interval. It is the confidence interval for the mean of the simple return relative to the S&P 500 of positive recommendations in the Featured section of *Mad Money*, where the stock is bought on the fifth trading day after the broadcast and held for 25 days. All confidence intervals, unless otherwise noted, were found via the statistical bootstrap (Section 5).

People have found incredibly clever ways to misinterpret confidence intervals. Figure 5 shows the bootstrap distribution of the mean of the data along with an indication of the confidence interval given in Table 1. The vertical line indicates the mean of the data. The distribution can be thought of as the probability of where the true mean lies.

Figure 5: Bootstrap distribution with 95% confidence interval of the mean simple return relative to the S&P 500 for Featured positive recommendations bought on day 5 and held 25 days. (TheStreet data)



First off, what do I mean by "true mean"? Our data are 505 instances of recommending a stock. We aren't really interested in those 505—we are interested in future recommendations. There is a population of future recommendations, and we are hoping the 505 recommendations that we have in hand are representative of that population. The task we have set ourselves is to learn the location of the mean of that population of future recommendations.

One misinterpretation of confidence intervals is that the truth is equally likely throughout the interval. That is virtually always wrong, and in particular it is wrong here. Values in the middle of the interval are much more likely than values near either edge.

There are two interpretations of confidence intervals. The first interpretation supposes that we generate a large number of intervals. In this interpretation the true value is either in the interval or outside of it—we don't know which. But if we generate 1000 confidence intervals (perhaps on 1000 different market commentators), then we expect that the true value will be in about 950 of the intervals. This is the frequentist interpretation.

The other interpretation is Bayesian. Bayesians will agree with the frequentist interpretation, but say there is no need to make it so complicated. A Bayesian is willing to say that the true value has a 95% probability of being in the interval we created. The key difference is the meaning of probability—a topic that is beyond the scope of this paper. Way, way beyond.

The CNBC analysis did not include any indication of variability. Table 2

Table 2: 95% confidence intervals for the mean simple return. (CNBC data)

holding period	n. obs.	mean	lower bound	upper bound
1 month	445	1.18%	0.47%	1.90%
2 months	445	2.50%	1.43%	3.60%

shows the 95% confidence intervals for these data. The lower bounds are above the S&P returns, but probably most people would put less weight on the data once they are given the confidence intervals. This is a particular instance of a seemingly large difference not being so large.

4.2.2 Adjusting for the Market

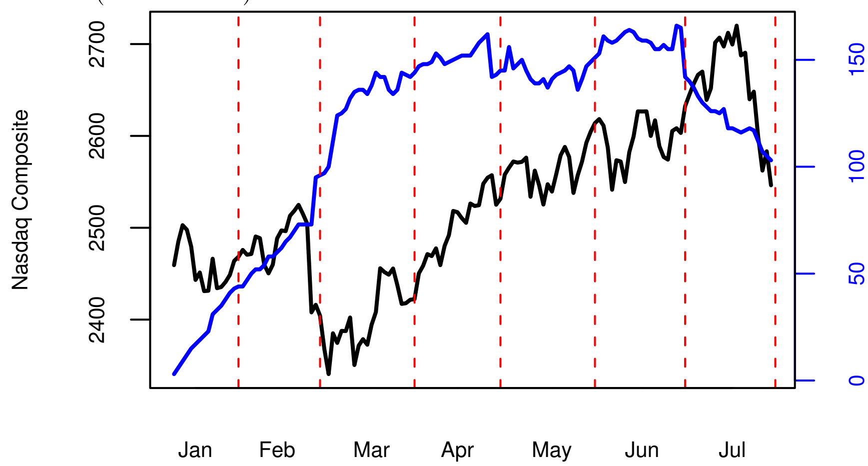
The Nayda and *Barron's* analyses adjust for the market day by day. In contrast the CNBC analysis does not. It merely compares the market return for a similar period of time. This implicitly assumes that the recommendations are evenly spread throughout the period of time quoted for the market. Figure 6[1] shows that not to be true, and that in this case there is significant bias.

When the number of recommendations is at full strength—mid March through June—the market steadily rose. The two points at which there were significant falls in the market were times when there were many fewer recommendations in place. Of course the bias could have easily gone the other way. Not adjusting for the market day by day for each recommendation adds a lot of noise to an already very noisy process.

[1]The R code to produce Figure 6 is:

```
function (filename='countlev.eps')
{
   if(length(filename)) {
      postscript(file=filename, height=3, width=5, horiz=FALSE)
      par(mar=c(5,4,0,4)+.1, cex.axis=.7, cex.lab=.7)
   }
   datenam <- names(count.cnbc2c)
   plot(nasdlev[datenam], type="l", axes=FALSE,
      xlab="", ylab="Nasdaq Composite", lwd=2)
   axis(2)
   par(new=TRUE)
   plot(count.cnbc2c, type="l", axes=FALSE, col="blue", xlab="",
      ylab="", lwd=2)
   axis(4, col="blue", col.axis="blue")
   mondiff <- c(0, diff(as.numeric(substring(datenam, 6, 7))), 1)
   whichdiff <- which(mondiff == 1)
   abline(v=whichdiff, col="red", lty=2)
   axis(1, at=whichdiff - 10, month.abb[1:7], tck=0)
   box()
   if(length(filename)) dev.off()
}
```

Figure 6: CNBC recommendation counts compared to the Nasdaq Composite. The black line (left scale) is the level of the Nasdaq Composite index; the blue line (right scale) is the number of stocks in the two-month window at each point in time. (CNBC data)



4.2.3 Returns

The Nayda analysis made a schoolboy error of summing simple daily returns to get longer period returns. Well, fair enough—he *is* a schoolboy, a fifth-grade schoolboy. This looks to be an error in favor of Cramer—the rejection of skill would probably be slightly stronger with the correct calculation.

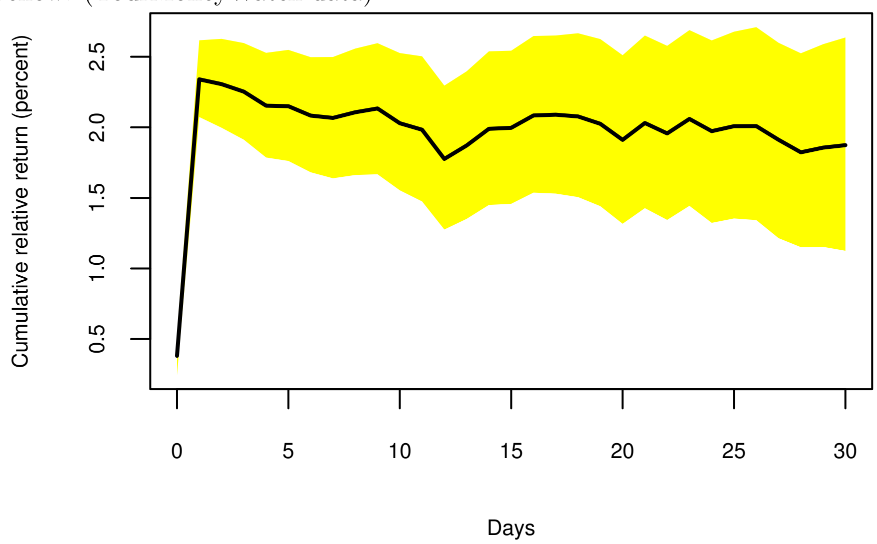
Even though summing simple returns across time is incorrect, it is not so uncommon for it to be done. Plots of cumulative simple returns are used. They are wrong, but not so wrong as not to be useful—especially when the purpose is to highlight differences of two strategies.

In our analyses should we use simple returns or log returns? That is a matter of opinion and circumstance.

Let's take a simple example of two stocks: one stock doubles in price, and the other halves in price. If we start off with one dollar in each stock, we end up with \$2.50, and hence a simple return of 25%. The average of the log returns is zero in this case. As suggested in Section 2 the weighted average of the simple returns produces the portfolio simple return. This argues for using simple returns in the analyses.

If instead of holding both stocks we are only going to hold one of them, the situation changes. Now we are either going to double our money or lose half of it (and we don't know which when we make our decision). Almost all people

Figure 7: The cumulative mean simple return from the day of the broadcast relative to the S&P 500. The pointwise 95% confidence interval is indicated in yellow. (YourMoneyWatch data)



are risk averse—losing an amount counts more against than gaining the same amount counts for. If there will be a selection of recommendations, then using log returns may be more appropriate.

Another factor is the statistical principle of being conservative. Simple returns make the buy recommendations look better, log returns make those recommendations look worse. Conservative analyses would mean that the CNBC analysis should have used log returns, and—given their generally insignificant results—the *Barron's* analyses should have used simple returns. In fact just the reverse was the case. Only the Nayda analysis was conservative—the conclusion was of no skill using simple returns.

That the *Barron's* analyses used log returns was due to my habitual behavior. My usual task is to prove that a strategy outperforms. The conservative approach in such cases is to use log returns. I didn't shift gears appropriately for the *Barron's* analyses once we saw indications of insignificance.

I also assumed that the difference when switching return types would be small. In fact it has a bigger impact than I supposed. Figure 7 can be compared to Figure 4 to see the effect of changing between simple and log returns.

4.2.4 Bias

It is always a good idea to ask if there are sources of bias in analyses. When working with returns, seemingly very small biases can have big impacts.

Table 3: 95% confidence intervals of the mean simple return relative to the S&P 500 using either all the data or only the relevant data. (CNBC data)

data	holding period	n. obs.	mean	lower bound	upper bound
full	1 month	445	0.26%	-0.39%	0.91%
full	2 months	445	1.11%	0.11%	2.14%
relevant	1 month	410	0.38%	-0.32%	1.08%
relevant	2 months	348	1.19%	0.01%	2.41%

Survival bias is popular in studies of equities. Suppose that we want to test a strategy on the S&P 500; so we get a ten-year history of the returns of the current constituents of the index. Using our strategy on this data, we compare the strategy results to the S&P 500. This is an example of survival bias—we have ruled out all of the stocks that fell out of the index, including those that went bankrupt. We have also selected only the stocks that entered the index and are still in it. The test is giving the strategy a big advantage.

In the present case of the *Mad Money* data there is only one instance of bias that I'm aware of. The CNBC data purportedly hold the recommendations for one or two months. In fact several of the recommendations were not old enough at the time the data were created to allow the full holding period. Some were only held for a day. The effect is to bias results towards zero. Assuming positive performance relative to the index, this is a bias against Cramer. Table 3 shows the extent of the bias.

4.2.5 What is the Question?

An analysis should be tailored to the question that it is trying to answer.

Tables 1, 2 and 3 give statistics on the mean return relative to the S&P for a specific strategy. Looking at the mean is appropriate if we are thinking about an individual who performs this strategy with many of the recommendations. That may be the most reasonable scenario. But suppose we are thinking of a large number of viewers each selecting one recommendation, looking at the median would then make sense. The median would speak to the typical viewer who implements a recommendation—half would do worse and half would do better. Table 4 presents the 95% confidence interval for the median of positive Featured recommendations when buying on the fifth day after broadcast and holding 25 days.

Comparing Figure 3 with Figure 4, it seems that the *Mad Money* recommendations have considerably more momentum than the S&P 500. This is circumstantial evidence that the Nasdaq Composite may be a more appropriate reference. Table 5 shows the difference that switching between these two indices makes.

Table 4: 95% confidence interval of the median simple return relative to the S&P 500 for Featured positive recommendations bought on day 5 and held 25 days. (TheStreet data)

n. obs.	median	lower bound	upper bound
505	0.36%	-0.27%	1.08%

Table 5: 95% confidence intervals of the mean simple return relative to either the S&P 500 or the Nasdaq Composite. (CNBC data)

holding period	index	n. obs.	mean	lower bound	upper bound
1 month	S&P 500	410	0.38%	-0.32%	1.08%
2 months	S&P 500	348	1.19%	0.01%	2.41%
1 month	Nasdaq	410	-0.01%	-0.70%	0.70%
2 months	Nasdaq	348	0.64%	-0.53%	1.83%

4.2.6 Trading Costs

Up to this point we have ignored trading costs. Quite the elephant to ignore. Suppose that we accept the 2.5% mean return of Table 2—the largest that we've seen (and it was not relative to an index). The cheapest trading that I see for an individual at the time of writing is US$7. In order to take advantage of a recommendation, a viewer needs to buy the recommendation, hold it, and then sell it. So there are $14 in trading costs. (There is also the bid-ask spread to consider, but that will almost always be trivial in comparison.) In order to break even, our viewer needs to buy $560 of each stock. If the true value of the recommendations is less than 2.5% or trading cost is higher, then the viewer will need to buy in even bigger blocks to break even.

The longest analyses we've discussed have two-month holding periods. This is a very short time frame given the costs incurred by non-professional traders. Longer time frames have a much better chance of paying off.

5 Bootstrapping

The statistical bootstrap is a means of assessing the variability of a statistic. It was introduced by [Efron, 1979]. Computers are now literally thousands of times better than in 1979, yet bootstrapping and its relatives are still not as widely used as they might be.

Figure 5 shows an example of a bootstrap distribution. The mean of 505 returns was found. We then want to know how variable that mean is. The idea of the bootstrap is that while our data consists of 505 specific numbers, we might

Table 6: Bootstrap and t distribution 95% confidence intervals of the mean simple return relative to the S&P 500 for Featured positive recommendations bought on day 5 and held 25 days. (TheStreet data)

method	n. obs.	mean	lower bound	upper bound
bootstrap	505	0.74%	-0.06%	1.60%
t distribution	505	0.74%	-0.09%	1.58%

have got some other set of 505 numbers. We want to make minimal assumptions regarding the datasets that we might have got. The bootstrap's solution to this problem is to sample our 505 numbers 505 times with replacement. A bootstrap sample will contain multiple copies of some of the numbers and no copies of other numbers.

To get the bootstrap distribution of our mean statistic, we do the following steps:

1. Produce a bootstrap sample by sampling the 505 datapoints with replacement 505 times.

2. Calculate the mean of the bootstrap sample produced in step 1, and save it.

3. Repeat steps 1 and 2 a large number of times. Ten thousand bootstrap samples were used in the analyses of Section 4.

The old-fashioned approach to get a confidence interval for a mean is to assume a normal distribution and use Student's t distribution. Returns are not normally distributed, but the mean of 505 returns is going to be quite close to a normal, so we don't need to worry about the assumptions. Table 6 shows the bootstrap and t distribution confidence intervals for the analysis we are focusing on. The differences are very small and of no consequence.

So if there is no real difference in the intervals, why bother with the bootstrap? It actually took me less time (including computing time) to get the bootstrap interval than to get the traditional interval. Plus I felt more confident with the bootstrap that I wasn't making a stupid blunder. While an introductory statistics book will tell you how to produce a confidence interval for a mean, that same book is not going to tell you how to produce a confidence interval for a median or a 5% trimmed mean or the raft of other statistics that you might find useful. That book is not going to tell you whether or not the sampling distribution of your statistic in a particular instance is close to the normal or not.

The world of the bootstrap isn't entirely rosy. There are a number of refinements—such as bias correction—that get around problems with the bootstrap in certain situations. There are even situations where those refinements fail as well. However, a lot of the time we don't care about formal inference, we just want a notion of whether or not the statistic is massively too variable to tell us much. The bootstrap will often work for this purpose even if it gives a biased impression of variability. There is a large literature on the bootstrap—starting points might be [Davison and Hinkley, 1997, Efron and Tibshirani, 1993].

6 Pseudo-Cramer

All of the analyses presented in Section 4 fail miserably at what should arguably be the real question: Does Jim Cramer have stock picking skill? A secondary question is how to take advantage of the skill. But if there is no skill, then there is nothing to take advantage of.

In the analyses that were done, some set of returns were compared to an index—mostly the S&P 500. The recommendations look very little like an index. Performance relative to the S&P 500 is largely driven by the behavior of the few largest capitalization stocks. If those few giant stocks have relatively good returns, then it will be very hard to beat the index. If the giants do relatively poorly, then it will be easy to beat the index. [Burns, 2007b] describes this and other problems of using indices for performance assessment.

There is a way to accurately evaluate stock-picking skill. Consider the recommendations as a portfolio—stocks enter the portfolio as they are recommended, and they leave the portfolio when the recommendation changes. This portfolio will have a number of characteristics: the number of stocks, its volatility, its distribution of market capitalization, and so on. A large number of random portfolios can be generated that have the same characteristics as the Cramer portfolio. The return of the Cramer portfolio would then be compared to the distribution of returns from the pseudo-Cramer portfolios.

Random portfolios—also called Monte Carlo portfolios—satisfy some given set of constraints but are otherwise free to have any holding in the universe of assets. More details on random portfolios including their advantages for performance measurement are given in [Burns, 2007b] and [Burns, 2007a]. Let's look at the constraints most likely to be useful in the current situation.

6.1 Number of Names

Matching the number of names is an obvious choice. We can expect the behavior of portfolios to change as the number of assets changes.

When mimicking investment funds, it is common to give a constraint on the maximum weight. In this case the assets are likely to be equally weighted, or perhaps to have a weight for buys and a larger weight for strong buys. Matching the weight distribution of the recommendation portfolio is sensible.

Equal weighting is artificial, but justifiable when testing recommendations. They do not work when testing trading strategies.

6.2 Volatility

One reason to use the same number of names in the random portfolios as are in the recommendation portfolio is to try to match the volatility. It is a basic tenet of finance that more volatile portfolios should produce higher returns. That the recommendation portfolio is almost sure to have a higher volatility than the S&P 500 (or even the Nasdaq Composite) means that comparing the recommendations to these indices could be giving the recommendations a large advantage.

We can produce random portfolios that have a volatility that is in a small neighborhood of the volatility that the recommendation portfolio happens to have. So the recommendations are at neither advantage nor disadvantage. This step needs the estimation of a variance matrix of the returns of the universe of assets, a quite easy task given price histories of the assets.

6.3 Market Capitalization

In the long run small capitalization stocks tend to outperform large cap stocks. However, there are periods when small caps strongly underperform large caps. Market capitalization is a large source of variation in returns that we can control for with random portfolios. We might partition the stocks into 10 or 20 categories of capitalization, and then select the same number of stocks from each category as are in the recommendation portfolio.

6.4 Other Constraints

Other sources of variation might be hypothesized. As we've seen there is an indication that Cramer tends to recommend high momentum stocks. The amount of momentum could be controlled for. Matching the sector allocation would be another possibility.

6.5 Implementation Details

So far the discussion of random portfolios has been as if the recommendation portfolio were static. In fact it will change—daily in Cramer's case. However, the key characteristics of the recommendation portfolio are unlikely to change dramatically. Hence the random portfolios need not be generated nearly so often as daily. However, they could be if a very strict control were desired.

The recommendation portfolio will have some return for a given time period. The random portfolios will produce a distribution of returns for that period. Unlike in the analyses of Section 4, here we can produce a rigorous test of skill. If there is quibbling about assumptions, then the set of constraints for the random portfolios can be adjusted. In any case, the assumptions are clearly visible.

The fraction of random portfolios with returns larger than the return of the recommendation portfolio is the p-value of a statistical hypothesis test of skill.

The null hypothesis is that there is no skill (certainly the random portfolios do not exhibit skill); the alternative hypothesis is that there is skill. This test is independent of whether we use simple returns or log returns. We can also test for negative skill.

7 Conclusion

In my opinion, Jim Cramer's stock-picking superiority is at best unproved. But being skeptical is my job. However, Jim Cramer is **not** on my personal list of most suspicious market commentators. That kudo would go to chartists. I find it hard to believe that anyone can tell what is going to happen to a price series in the future by looking at the pattern of its history. On the other hand, I'd love to be proved wrong. A database of picks that are tested against random portfolios would be the most likely way of proving me wrong.

All media commentators that issue specific recommendations should maintain a database so that tests of skill as outlined here would be easy to perform. Analysts to perform the tests are also needed—the work involved is mainly data gathering. If these two elements were in place, investors would be well served.

References

[Burns, 2007a] Burns, P. (2007a). Bullseye. *Professional Investor.*

[Burns, 2007b] Burns, P. (2007b). Random portfolios for performance measurement. In Kontoghiorghes, E. J. and Gatu, C., editors, *Optimisation, Econometric and Financial Analysis*. Springer.

[Davison and Hinkley, 1997] Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

[Efron, 1979] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1–26.

[Efron and Tibshirani, 1993] Efron, B. and Tibshirani, R. (1993). *Introduction to the Bootstrap*. Chapman & Hall.

[Engelberg et al., 2007] Engelberg, J., Sasseville, C., and Williams, J. (2007). Attention and asset prices: the case of Mad Money. Technical report, SSRN, http://papers.ssrn.com.

[Nayda, 2006] Nayda, N. (2006). You! me! Let's get ready to make Mad Money!!! Science fair project, Elmont Virginia Elementary School.

[R Development Core Team, 2007] R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, http://www.r-project.org. ISBN 3-900051-07-0.