

Questions.

1. Calculate the annual and daily growth and volatility estimates for the DJIA and 4 stocks comprising the ISCO index. Note that the daily growth values will be pretty small.

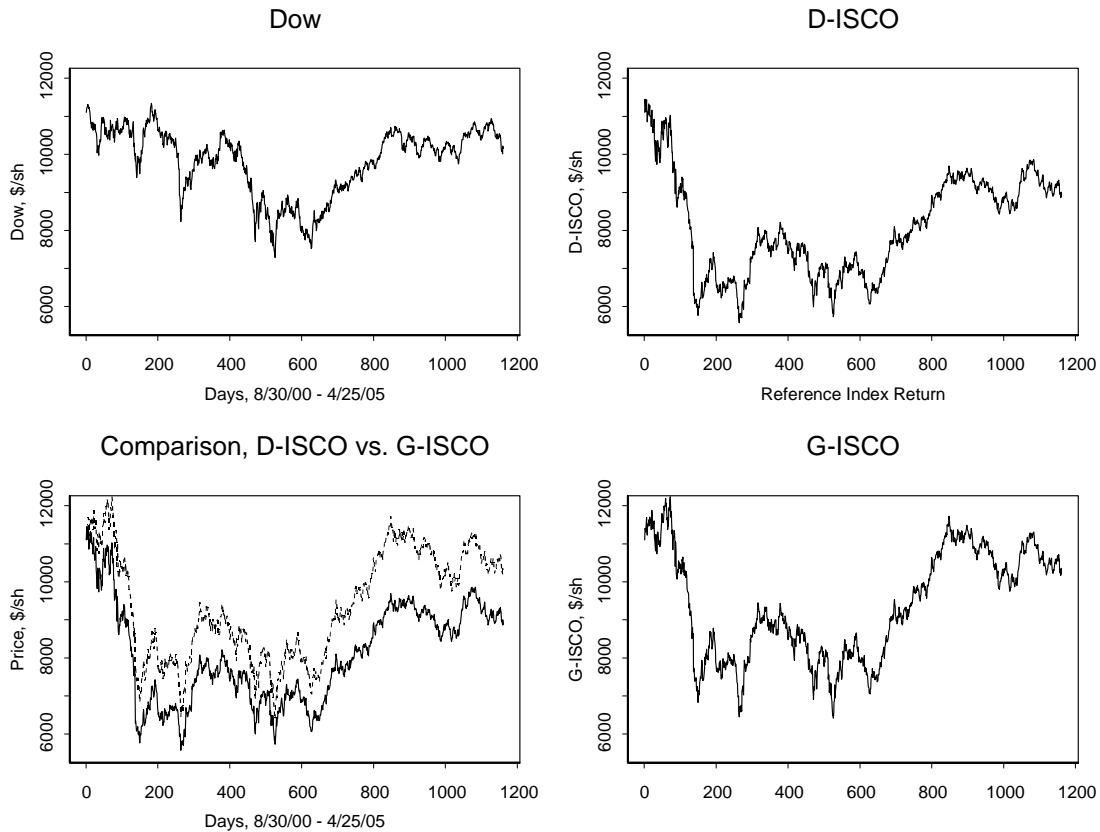
mu_daily	-0.000029	-0.00050379	0.000566	0.001084	-0.000006
sigma_daily	0.030123	0.03609181	0.016251	0.018321	0.011933
mu_annual	-0.007425	-0.12695413	0.142691	0.273047	-0.00139
sigma_annua	l 0.4781885	0.57293973	0.257984	0.290842	0.189423
r_bar	-0.0004832	-0.0011551	0.000434	0.000916	-0.000077
sigma	0.030123	0.03609181	0.016251	0.018321	0.011933
CSC	CSCO	SY	SJM	DJIA	
8/30/00	76.19	66.56	20.97	19.78	11103

For convenience and reference, we also calculate parameters for the generated indexes:

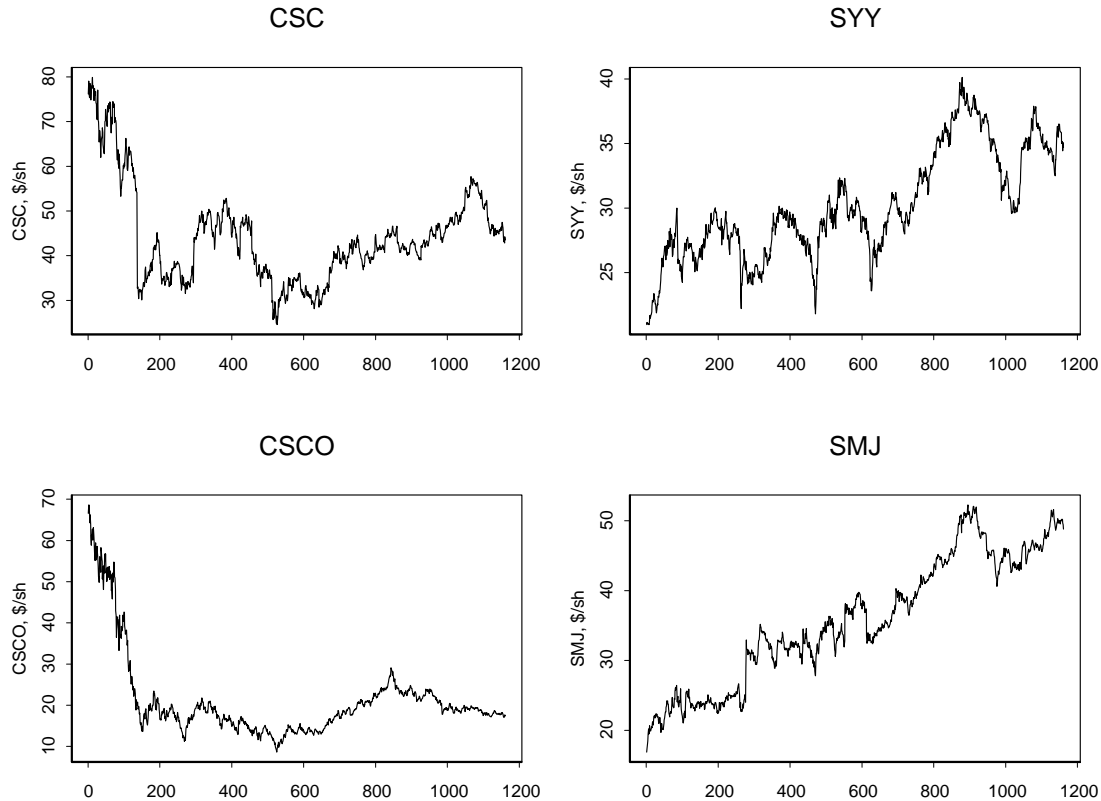
mu_daily	-0.000053	0.000063
sigma_daily	0.016735	0.016423
mu_annual	-0.0132496	0.015814
sigma_annual	0.2656672	0.260701
r_bar	-0.0001926	-0.000072
sigma	0.0167355	0.016423
D-ISCO	G-ISCO	
11103	11103	
First Value	45.15	36.60
Multiplier	245.92724	303.3780

2. Plot price charts for the DJIA and each component stock.

First we show the Dow as compared with the ISCO indexes. Note that the scale plotted is the same for all graphs, enabling one to immediately see that the Dow held up better than ISCO. A little harder to see is that the G-Isco got higher than the Dow ever did (around day 70).



The ISCO stock components are shown below:



3. Perform a correlation and regression analysis for each ISCO stock and the DJIA. Try to make statements on what the correlations mean. Plot your regression lines.

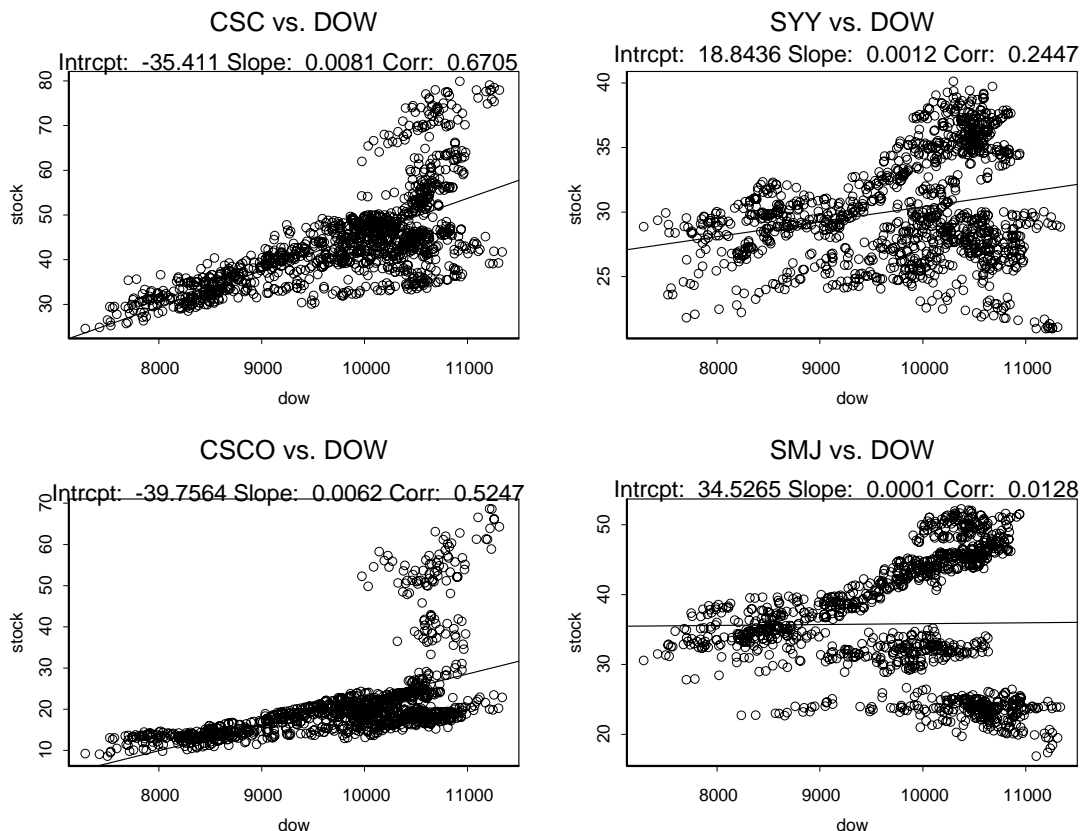
We provide what is known as the "correlation matrix", which is symmetric, with $\text{Corr}(X_{ij}) = \text{Corr}(X_{ji})$. On the diagonals are the correlation of each index/stock with itself (1.00); the intersection of any row/column (X_{ij}) is the correlation between that pair.

Correlation Analysis

	DOW	Disco	Gisco	CSC	CSCO	SY	SMJ
DOW	1	0.654	0.669	0.671	0.525	0.245	0.013
Disco	0.654	1	0.981	0.811	0.732	0.399	0.341
Gisco	0.669	0.981	1	0.721	0.646	0.52	0.446
CSC	0.671	0.811	0.721	1	0.833	-0.125	-0.194
CSCO	0.525	0.732	0.646	0.833	1	-0.22	-0.329
SY	0.245	0.399	0.52	-0.125	-0.22	1	0.866
SMJ	0.013	0.341	0.446	-0.194	-0.329	0.866	1

Some discussion of the correlations in the above table follows. We note that all the stocks are positively correlated with the DOW; CSC and CSCO moderately-high correlated ($\sim .50$ -.60), and SYV/SMJ are faintly correlated. The correlation between CSC and CSCO is high, almost 84%, and SYV and SMJ are also high at 87%. Both pairs of stocks are in the same sector (tech and food service), so this is not surprising. Both SYV and SMJ are negatively correlated with CSCO/CSC, which is good in portfolio theory.

Notice that the regression line slopes are very small, but that is only because the Indexes themselves are so large. If the index changes by +1 point, the stock changes by +.008 point.

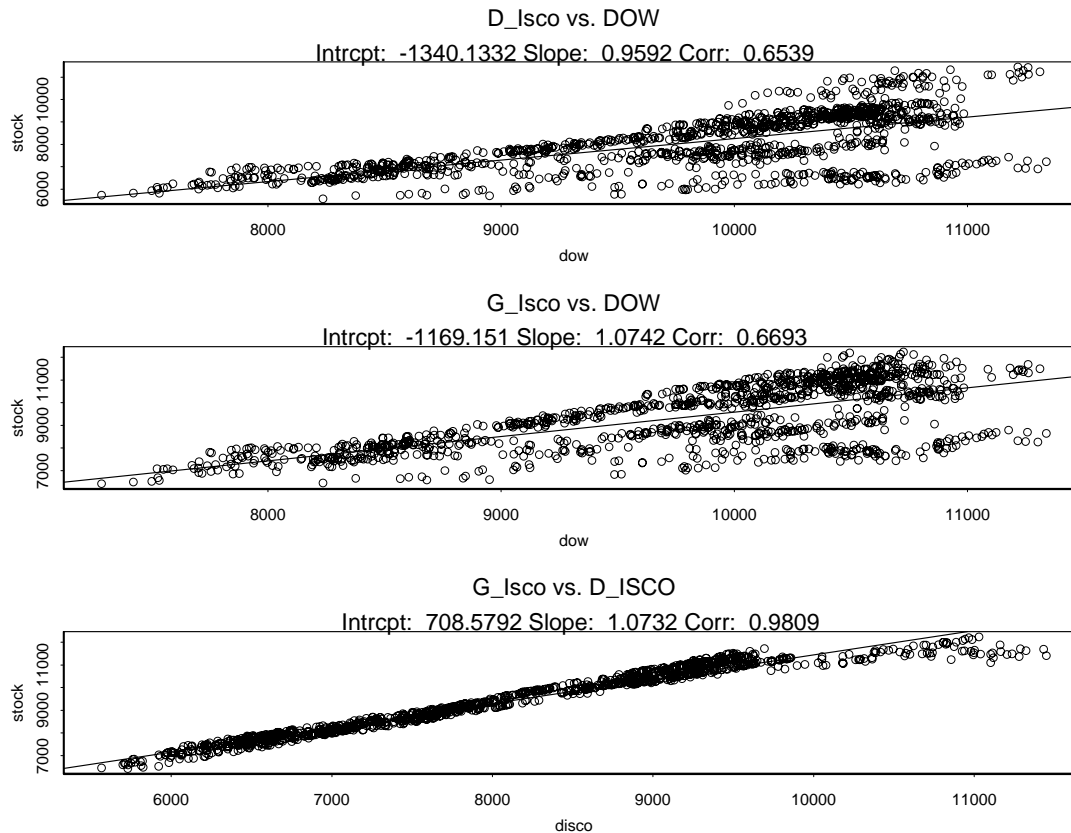


These are correlations between the indexes. Note that the ISCO indexes are only about 70% correlated with the DOW, indicating a possible portfolio diversification benefit.

	DOW	Disco	Gisco
DOW	1	0.654	0.669
Disco	0.654	1	0.981
Gisco	0.669	0.981	1

The DISCO/GISCO correlation is almost 98%, which is not surprising since they are composed of exactly the same stocks.

We can also look at the regressions among the indexes (below). Note here that the regression line slope is much larger and easier to get a feel for.



4. Calculate and plot the D-ISCO and G-ISCO indexes, scaling to the value of the DJIA on 8/30/2000.

See plots for problem 1

5. Find the correlation between the 4 component stocks and each of the ISCO indexes. Comment on this correlation with respect to those for the DJIA.

This is the same table as above:

	DOW	Disco	Gisco	CSC	CSCO	SYT	SMJ
DOW	1	0.654	0.669	0.671	0.525	0.245	0.013
Disco	0.654	1	0.981	0.811	0.732	0.399	0.341
Gisco	0.669	0.981	1	0.721	0.646	0.52	0.446
CSC	0.671	0.811	0.721	1	0.833	-0.125	-0.194
CSCO	0.525	0.732	0.646	0.833	1	-0.22	-0.329
SYT	0.245	0.399	0.52	-0.125	-0.22	1	0.866
SMJ	0.013	0.341	0.446	-0.194	-0.329	0.866	1

6. Would it be better to invest in the ISCO index or the DJIA? Why?

There are many ways to answer this question. They would fall into an analysis of the "buy-and-hold (BAH)," "portfolio re-balancing", or "trading" strategies. We can measure their performance over a horizon of 5 years, assuming an initial investment of say, \$100,000.

For the BAH strategy, the D-ISCO index UNDERPERFORMED the DOW, and the G-ISCO index slightly beat it.

	Initial value 100,000						
	8/30/00	8/31/01	8/30/02	8/29/03	8/31/04	4/8/05	Terminal Portfolio Value
							BAH
CSC	76.19	37.6	36.83	42.57	46.35	45.62	59,877
CSCO	66.56	16.33	13.82	19.14	18.76	17.9	26,893
SYN	20.97	28.02	28.36	30.64	31.74	35.99	171,626
SMJ	16.87	25.96	34.42	39.23	45.5	49.67	294,428
DJIA	11,103	9,950	8,664	9,416	10,174	10,461	94,221
DISCO	11,103	6,635	6,974	8,090	8,752	9,172	82,607
GISCO	11,103	7,843	8,055	9,544	10,156	10,545	94,976

Even with the DOW, however, you lose money over a really long time (4 years). Not counting inflation, which is a lot more than the 2.5% that is reported, you lose 6% on the DOW, 18% on the DISCO, and about 5% on the GISCO. What drives the game is noticing that "Hey, if I'd invested in SJM I would have tripled my money!"

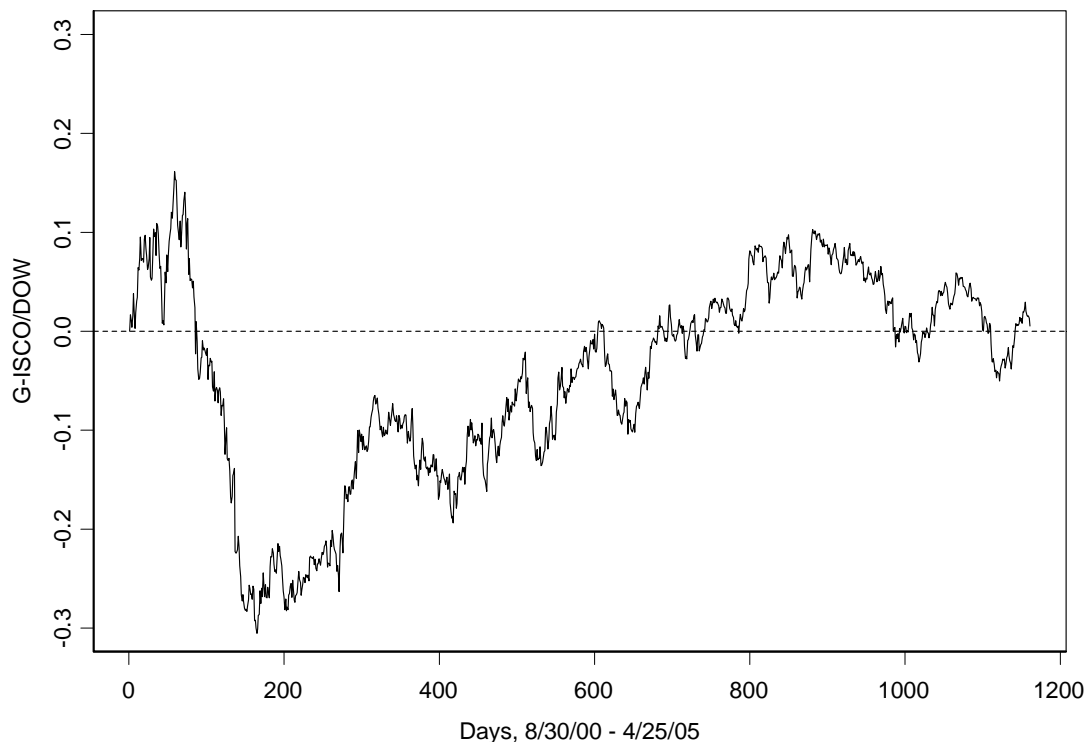
The portfolio re-balancing or trading strategy would show a different story. Since the markets started a major decline (based on technical performance) starting March 2000, you would be careful when investing huge chunks of money. Suppose you had a protection scheme in place which sold you out after your portfolio showed a slight loss after 8/2000; you stayed out until 8/02, and then re-entered on 9/3/02 (again, based on your crystal ball). Then we would have the following returns:

	Initial value 100,000						
	Buy Here				Terminal Portfolio Value		
	9/3/02	8/29/03	8/31/04	4/8/05	BAH	Buy in 02	
CSC	35	42.57	46.35	45.62	59,877	130,343	
CSCO	13.07	19.14	18.76	17.9	26,893	136,955	
SYN	27.77	30.64	31.74	35.99	171,626	129,600	
SMJ	34	39.23	45.5	49.67	294,428	146,088	
DJIA	8,308	9,416	10,174	10,461	94,221	125,918	
DISCO	6,753	8,090	8,752	9,172	82,607	135,816	
GISCO	7,777	9,544	10,156	10,545	94,976	135,588	

Here both ISCO indexes did much BETTER than the DOW. Even if you suffered a 5% loss in 2000, you'd still be way ahead after staying out for 2000 to mid-2002, and either ISCO investment would have been much better.

Another way to compare relative performance would be to look at the difference between the indexes (such as $D - ISCO$) each day. In statistics and mathematics, this usually is interpreted as a ratio, like $G - ISCO / DOW$, which begins at 1.00 on day 1, or 0.00 if you subtract 1.0. This appears as:

Comparison, G-ISCO vs. DOW



For the BAH strategy, if you could buy the ratio, it is apparent that the Isco index immediately outperforms, then quickly underperforms the Dow, and that you would have to wait 700 days or so just to break even. And then, you would be disappointed by the end of the holding period when Isco closed flat at the end of the period.

For a trading approach, the comparison chart clearly shows that the BEST time to sell the DOW and go long the ISCO would be around day 180, or 8 ½ months after the 8/00 start date.

Unfortunately, in the past you don't know the future. Using technical analysis, you would have waited for the pop up around day 190, wait for it to settle back down (around day 200); when it did not go below the previous low, you would have gone long

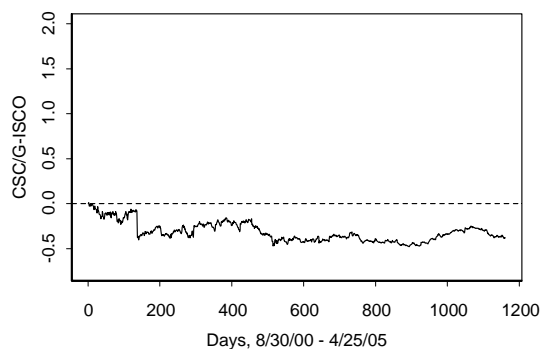
around day 275. Note that the market was aware of all this; do you see how at day 275 the big jump over the top from day 199 and 250? So "everyone" was getting in on this signal.

Sure enough, after a rise, the ratio declines again to a low around day 410, which would have made many people sell the ratio. If you stayed in, however, buying on day 300 (November 14, 2001) would have handsomely rewarded you.

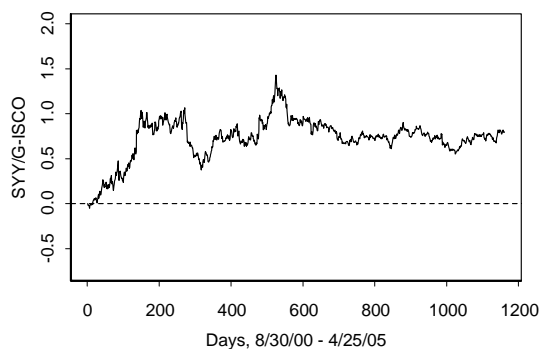
7. Would it be better to invest in the ISCO index or for an individual stock in the ISCO index? Why?

This one is a little trickier, since you would have to know which Isco stock would outperform the others. There is no way to tell that SYJ/SMJ would do better than CSC/CSCO, other than a prediction that the "tech-wreck" which began March 2000 would continue for 4 years. There are divergence measures you could employ to tell which sector was doing better, in which case it would have been better to reduce your allocation of the tech components and increase those for the food service stocks. This is very hard to do, however, and most advisors fail.

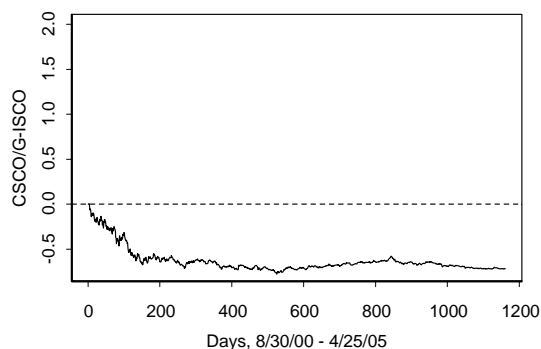
Comparison, Stocks to G-ISCO



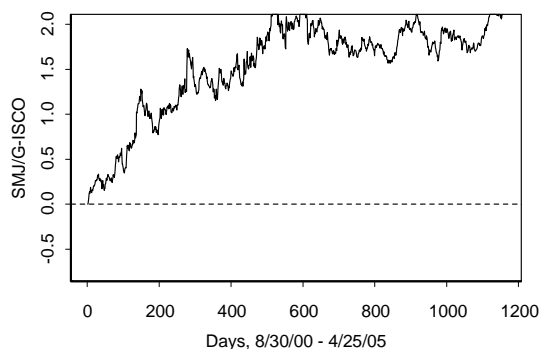
Comparison, Stocks to G-ISCO



Comparison, Stocks to G-ISCO



Comparison, Stocks to G-ISCO



8. Suppose after the Bush re-election each of the ISCO stocks gained 10% return on the day. What are the p-values for these returns how often should such a return be encountered in terms of years (or centuries, or millennia)?

Recall the daily parameters for our 4 stocks:

	CSC	CSCO	SY	SJM	DJIA	DISCO	GISCO
mu_daily	-0.00003	-0.0005	0.000566233	0.001084	-5.5E-06	-0.000053	0.000063
sigma_daily	0.03012	0.036092	0.01625146	0.018321	0.011933	0.016735	0.016423
R	1.10	1.10	1.10	1.10	1.10	1.10	1.10
r = ln(R)	0.095	0.095	0.095	0.095	0.095	0.095	0.095
$z = \frac{r - \mu_{daily}}{\sigma_{daily}}$	3.165	2.655	5.830	5.143	7.988	5.698	5.800
p-val	0.00155	0.007937	5.56458E-09	2.71E-07	1.33E-15	1.21E-08	6.66E-09
days	644.8	126.0	179,708,186	3,691,884	7.51E+14	82374876	1.5E+08
years	2.56	0.50	713,128	14,650	2.98E+12	326,884	595,693

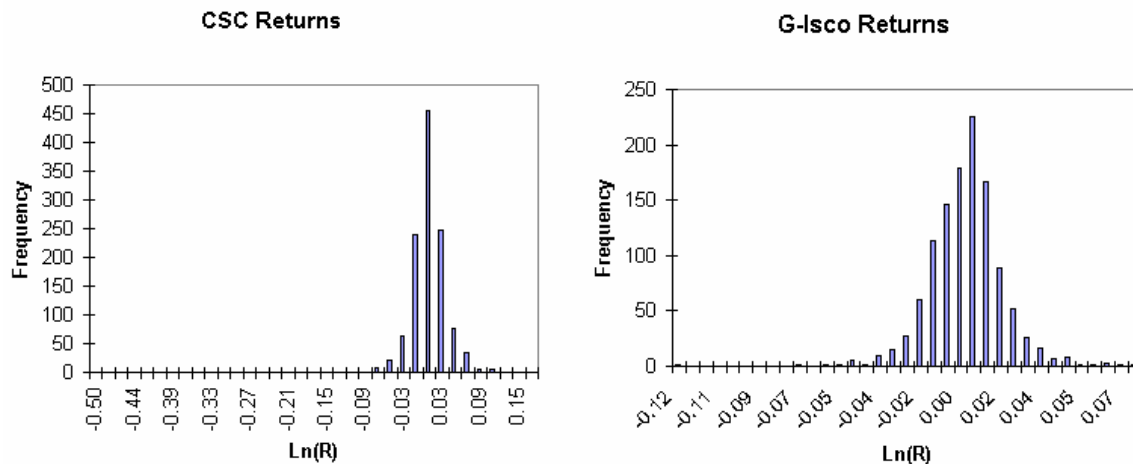
A 10% move on a stock with a daily volatility of 1% is a "10-sigma" event; a 10% move on a stock such as CSCO with an almost 4% daily volatility is only 2.5 sigma, which should occur about 3 times a year, based on the p-values and time-between-events indicated in the table.

9. Do your results in question 8 make you wonder about the Normality assumptions of the log of the daily returns?

a. Since we have so many greater than 6σ events happening so often, and since the Normal distribution just hardly ever gives values over 4σ , no matter how long you wait, it begs the question, "how can these returns be normal?" Much research has gone into this, and many solutions are attempted, mostly exotic, fat-tailed continuous distribution models, some of which do not even have finite moments such as the mean or standard deviation. Other models are employed which include a time-varying volatility concept.

b. Numerical analysis of the index or stock returns reveals many departures from normal, depending on how much effort you go to find them. One could calculate z-scores on all the returns and sort them, counting the number that are greater than 3. As an example of the latter approach, we have at least 15 that are greater than 3σ in absolute value, when based on our sample size we should only have about 4. You could examine the higher moments (skew, kurtosis, etc.), and observe that departures from normality are significant. You could perform a

histogram analysis on the returns, and observe the left-skewedness (and slightly so to the right), and conclude they are not quite from a normal population (see below).



c. More simply, after calculating the 4 returns in part 8, since we have so many greater than 3σ events happening, and since the Normal distribution just hardly ever gives

Check out these "humdingers" (you can do this daily using <http://fast2.quote.com/fq/stocktrak/group?mode=IwatchMenu1&group=IndustryWatch/ga&mode=ga&page=stocks>)

Wednesday, December 15, 2004 16:02 PM Eastern Time

Symbol	Name	Last	Change	Change%	Volume
LVS	NYSE:LVS (IPO)	46.56	17.56	60.55	26,260,300
CKCM	Click Commerce, Inc.	13.13	2.29	21.13	795,867
OPWV	Openwave Systems Inc	16.11	1.89	13.29	5,940,279
TRGL	Toreador Resources Corporation	18.73	2.16	13.03	451,379
LEN	Lennar Cp Cl A	56.35	5.52	10.85	9,374,600
LEN.B	Lennar Corp Cl B	51.85	4.97	10.6	580,000
MSO	Martha Stewart Liv	30.05	2.81	10.31	2,627,000

We see several 10% returns on this and almost every day; are they significant? I just checked three, and here are the parameters and the z-scores:

```
ckcm00-04    Mu_hat =    0.58021021    Sigma_hat =    1.27300990
opwv99-04    Mu_hat =    0.47903508    Sigma_hat =    1.21799147
len00-04     Mu_hat =    0.40983891    Sigma_hat =    0.41790226
```

Clearly CKCM and OPWV have volatilities over 120% per year! So these moves might not be that large. For Lennar homes, with 42% volatility, the 10% move might be unusual.

```

CKCM  mu_1d: 0.0023    sd_1d: 0.0802    z: 2.359  p_val:0.92E-02 Yrs: 0.433
OPWV  mu_1d: 0.0019    sd_1d: 0.0767    z: 1.602  p_val:0.55E-01 Yrs: 0.072
LEN   mu_1d: 0.0016    sd_1d: 0.0263    z: 3.854  p_val:0.58E-04 Yrs: 68.4

```

Stocks too low priced to really invest in; these show "non-normal" returns more often

Symbol	Name	Last Change	Change%	Volume
SQNM	Sequenom, Inc.	1.63	0.52	46.84 16,066,866
NAVI	NaviSite, Inc.	3.02	0.77	34.22 117,569
CDIC	CardioDynamics International Corporation	5.37	1.16	27.55 1,729,825
EVST	Everlast Worldwide Inc.	4.55	0.92	25.34 122,113
BWNG	Broadwing Corporation	8.48	1.48	21.14 5,039,583
LOUD	Loudeye Corporation	2.49	0.4	19.13 16,350,960
CVSN	ChromaVision Medical Systems, Inc.	1.69	0.22	14.96 788,151
HANA	Hanaro Telecom, Inc.	3.44	0.44	14.66 153,295
ATCO	American Technology Corporation	7.50	0.88	13.29 1,080,026
WGRD	WatchGuard Technologies, Inc.	4.45	0.48	12.09 1,041,819
ARIA	ARIAD Pharmaceuticals, Inc.	7.17	0.77	12.03 2,894,132
PYX	Playtex Products Inc	7.70	0.66	9.37 963,700

10. What other questions does your analysis bring to mind?

Notice that we did not attempt to perform any hypothesis testing in this lab. You could think of some tests. For example, are CSC and CSCO really any different? The Null hypothesis would assume the lognormal model, and that the parameters are the same. Your test statistics would be the difference. Under the null, you could clump together all the CSC and CSCO returns, draw two samples, calculate the parameters, subtract them, and calculate the probability that they exceed a certain value.

The first step is assessing the equality of the volatility parameters. Assuming a normal distribution of the returns, which on the surface appears ok, we use the F test for variance. Using the first 1000 days of CSC and CSCO returns, recalling that

$\sigma_{daily}^2 = Var(r)$, we obtain $\sigma_{csc o} = 0.0383$ and $\sigma_{csc} = 0.03183$, resulting in

our test statistic $F_{999,999} = \frac{\sigma_{csc o}^2}{\sigma_{csc}^2} = 1.450898$, which has a p-value of

Zero. We need to reduce the sample size to 80 in order to fail to reject at the 5% level. Although the variances appear very close, with a sample size of 1000 the difference is very significant, and we cannot pool the standard deviations. We would conclude (with little value) the means are equal, with

$$\sigma_{\Delta \bar{r}} = \sqrt{(\sigma_{csc o}^2 + \sigma_{csc}^2)/1000} = 0.00158 \quad \text{and} \quad z = \frac{\Delta \bar{r} - 0}{\sigma_{\Delta \bar{r}}} = \frac{-0.000773}{0.00158} = -0.49.$$

Remember, in this model, volatility is the key parameter.

There are other tests too, such as distributional tests, non-lognormal tests, etc. That would be a follow-on to this lab, but would be questions you might have.

Other questions might be:

- How can anyone make money in these situations?
- What is stealth correlation, and what does correlation mean in the context of financial markets, especially since the correlation changes all the time?
- How could we model a trading scheme and test it in time?
- What other statistical tools are required to understand financial markets?
- Since $\hat{X} \leq \bar{X}$, why is the scaled geoindex greater than the D-ISCO much of the time?
- Is there some sort of other model which could better explain the financial markets? We resort to stochastic models when a physical model is too complicated. Perhaps there is a better "physical" paradigm that would account for (and hopefully predict) these markets.
- How is it that stock market analysts make 6-figures when they are as clueless as the average monkey throwing darts at a page of mutual funds?
- Why do correlations based on *returns* differ from those based on *price*?

These and many other questions come to mind after performing the simple tests we have studied in the course.

The following are some questions raised by students in the past:

- o Is there a political regime effect on the markets? I.e., is there a significant return during periods of "pro-business" vs. "anti-business" political regimes?
- o "A better question would be, what's the probability of a massive market withdrawal from American markets in favor of unstable, developing nations due to a Bush reelection?"

- o What sorts of analyses are needed to correlate market activity with business cycles?
- o How does this fit into an overall seasonality of stocks/sectors/markets?
- o What impacts do "terrorist attacks" have on the markets in general and which sectors or markets do well in these events? This is known as "structural change" in the time series literature.
- o General probabilistic question: What is the probability that an individual stock increases and how does this relate to its GBM growth parameters?
- o Is there fruit in coefficient of variation analysis for stocks (σ/μ)?
- o What do the intercepts mean on the stock regressions, especially when they are negative, or if Dow = 4320 \rightarrow CSC = 0?
- o Why are the DISCO and GISCO not 100% correlated? Does it have something to do with the different type of averaging?
- o How does one measure "consumer confidence," and does it attempt to discover individual psychology?
- o Behavioral market efficiency: what causes investors to scare away from "undervalued" stocks when analysts are making strong buy recommendations?
- o Is there a statistically significant pattern on the scatterplots for SY/DOW and SJM/DOW? Sure looks like one. "I noticed some funny patterns in the scatterplots; it seemed to me it was likely due to difference scenarios at different times."
- o If one regresses the DOW on CSC, one sees a pattern which looks like could be brought linear via a simple transform. Explore this.
- o How develop a model which takes into account the expectations of the market (without modeling them as part of the EMH diffusion).
- o What other type of regression might be appropriate (besides linear)?

- o How should one test for non-normality of the market returns?
- o Since stocks trade "continuously", how do we justify a daily sampling given the Nyquist sampling theory?

Please feel free to call or write if you have any questions, former students always have unlimited technical support.