Foundations of Financial Markets
C15.0002
Summer 2006
Prof Yang Lu

Computing means, variances and covariances in Excel

In this mini-tutorial, we will illustrate how to use Excel to compute means, variances and covariances.

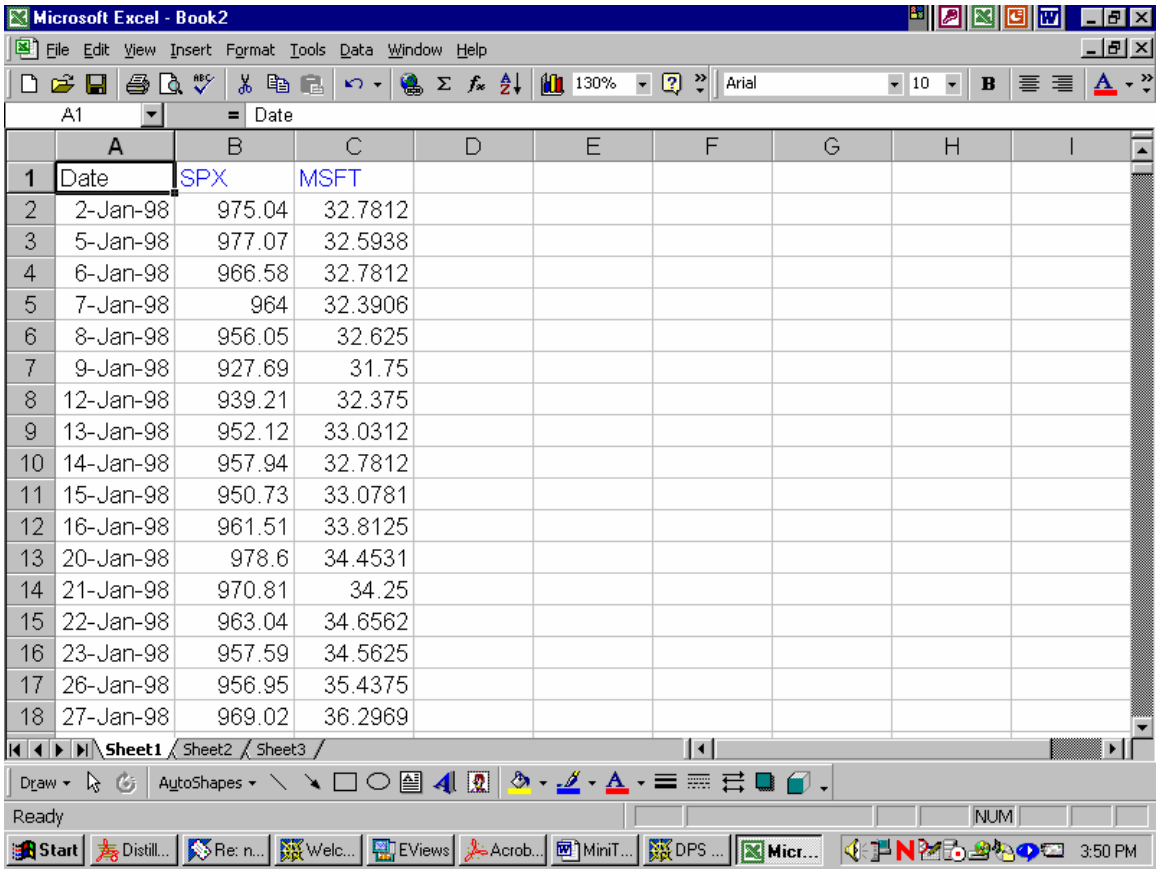
The **mean or (arithmetic) average** of a sequence of numbers $a_1, a_2, \ldots, a_n$ is defined as

$$m = \frac{1}{n}\sum_{i=1}^{n} a_i \quad .$$

In Excel, there is a built in formula to compute this quantity:

=AVERAGE(array),

where "array" denotes the array of cells to be averaged. For example, let's say that you are given a time series of daily closing prices on Microsoft stock and the S&P 500 Index for the period Jan 2nd, 1998 – March 6th, 2001. You first want to know the historical average return over this period.



	A	B	C
1	Date	SPX	MSFT
2	2-Jan-98	975.04	32.7812
3	5-Jan-98	977.07	32.5938
4	6-Jan-98	966.58	32.7812
5	7-Jan-98	964	32.3906
6	8-Jan-98	956.05	32.625
7	9-Jan-98	927.69	31.75
8	12-Jan-98	939.21	32.375
9	13-Jan-98	952.12	33.0312
10	14-Jan-98	957.94	32.7812
11	15-Jan-98	950.73	33.0781
12	16-Jan-98	961.51	33.8125
13	20-Jan-98	978.6	34.4531
14	21-Jan-98	970.81	34.25
15	22-Jan-98	963.04	34.6562
16	23-Jan-98	957.59	34.5625
17	26-Jan-98	956.95	35.4375
18	27-Jan-98	969.02	36.2969

The first thing to do is to get the daily rates of return, using the formula

$$R_t = (P_t - P_{t-1} + d_t) / P_{t-1}$$

where

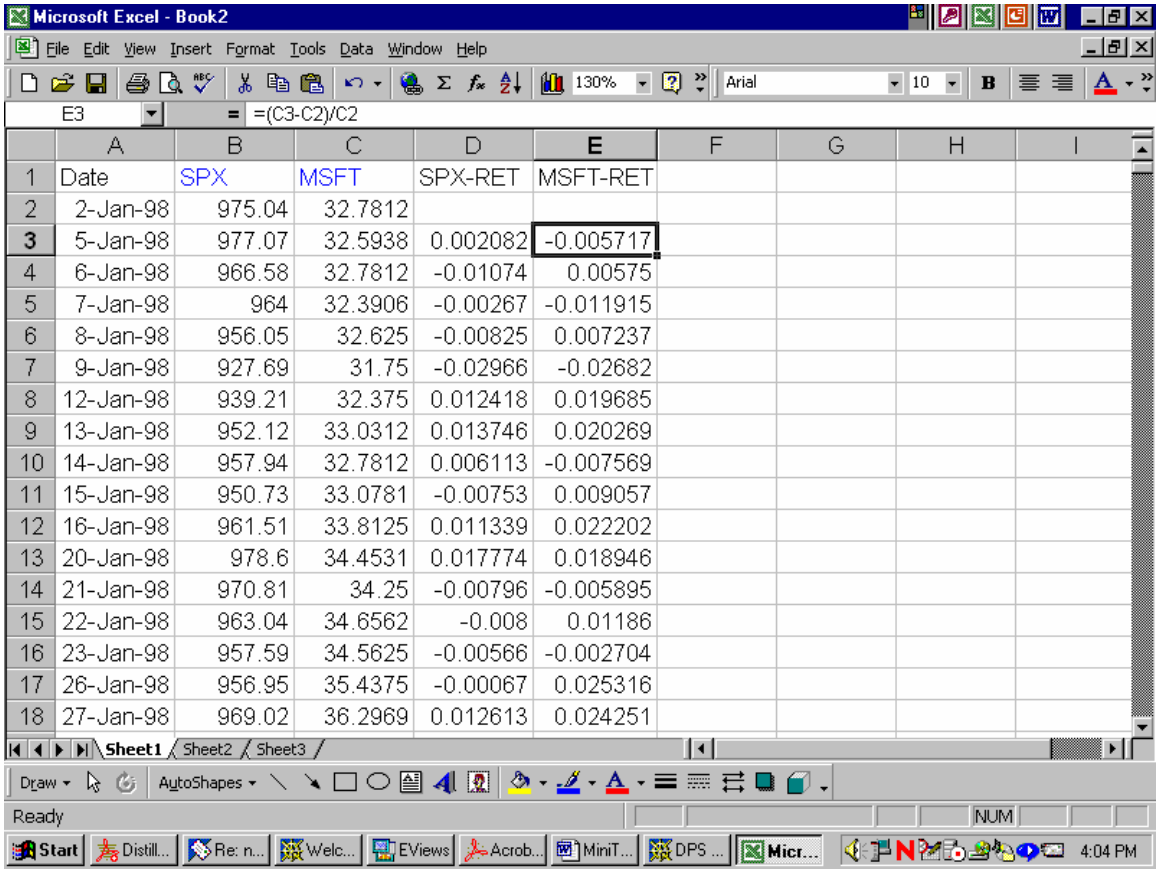
- R_t is the rate of return for day t,
- P_t is the closing price on day t,
- P_{t-1} is the closing price on day t-1,
- d_t is the dividend paid on day t.

Remark: You could also use the so-called continuously compounded rate (using logs); in practice the difference between the two is usually very small.

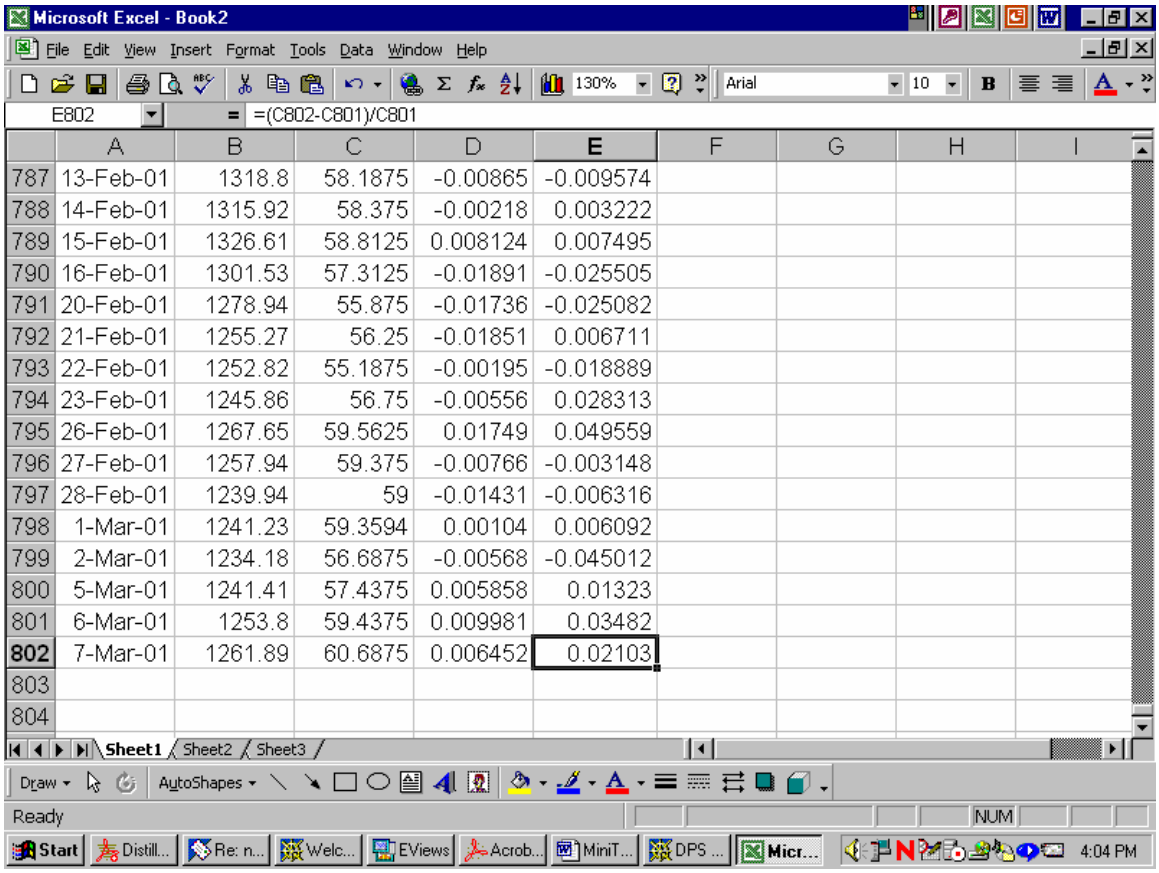
Microsoft does not pay any dividends, and we will ignore index dividends for simplicity. Therefore, type the following in cell D3:

=(B3-B2)/B2,

which is our first return on the S&P Index. You can drag this formula to the right to obtain the first return on Microsoft, and then down to get the entire time series.



	A	B	C	D	E	F	G	H	I
1	Date	SPX	MSFT	SPX-RET	MSFT-RET				
2	2-Jan-98	975.04	32.7812						
3	5-Jan-98	977.07	32.5938	0.002082	-0.005717				
4	6-Jan-98	966.58	32.7812	-0.01074	0.00575				
5	7-Jan-98	964	32.3906	-0.00267	-0.011915				
6	8-Jan-98	956.05	32.625	-0.00825	0.007237				
7	9-Jan-98	927.69	31.75	-0.02966	-0.02682				
8	12-Jan-98	939.21	32.375	0.012418	0.019685				
9	13-Jan-98	952.12	33.0312	0.013746	0.020269				
10	14-Jan-98	957.94	32.7812	0.006113	-0.007569				
11	15-Jan-98	950.73	33.0781	-0.00753	0.009057				
12	16-Jan-98	961.51	33.8125	0.011339	0.022202				
13	20-Jan-98	978.6	34.4531	0.017774	0.018946				
14	21-Jan-98	970.81	34.25	-0.00796	-0.005895				
15	22-Jan-98	963.04	34.6562	-0.008	0.01186				
16	23-Jan-98	957.59	34.5625	-0.00566	-0.002704				
17	26-Jan-98	956.95	35.4375	-0.00067	0.025316				
18	27-Jan-98	969.02	36.2969	0.012613	0.024251				



	A	B	C	D	E	F	G	H	I
787	13-Feb-01	1318.8	58.1875	-0.00865	-0.009574				
788	14-Feb-01	1315.92	58.375	-0.00218	0.003222				
789	15-Feb-01	1326.61	58.8125	0.008124	0.007495				
790	16-Feb-01	1301.53	57.3125	-0.01891	-0.025505				
791	20-Feb-01	1278.94	55.875	-0.01736	-0.025082				
792	21-Feb-01	1255.27	56.25	-0.01851	0.006711				
793	22-Feb-01	1252.82	55.1875	-0.00195	-0.018889				
794	23-Feb-01	1245.86	56.75	-0.00556	0.028313				
795	26-Feb-01	1267.65	59.5625	0.01749	0.049559				
796	27-Feb-01	1257.94	59.375	-0.00766	-0.003148				
797	28-Feb-01	1239.94	59	-0.01431	-0.006316				
798	1-Mar-01	1241.23	59.3594	0.00104	0.006092				
799	2-Mar-01	1234.18	56.6875	-0.00568	-0.045012				
800	5-Mar-01	1241.41	57.4375	0.005858	0.01323				
801	6-Mar-01	1253.8	59.4375	0.009981	0.03482				
802	7-Mar-01	1261.89	60.6875	0.006452	0.02103				
803									
804									

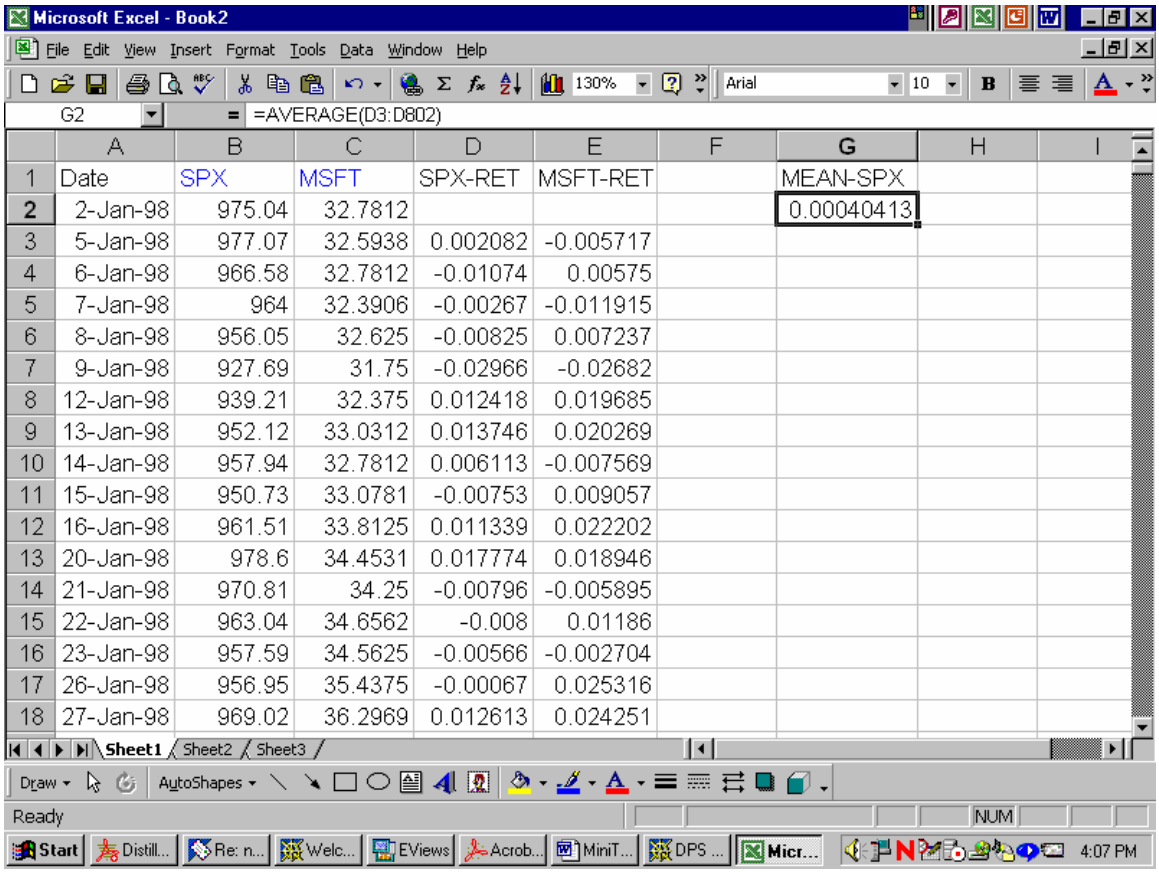
Now, to get the average return on the index, type (in any cell, say cell G2)

=average(D3:D802)

You can also type

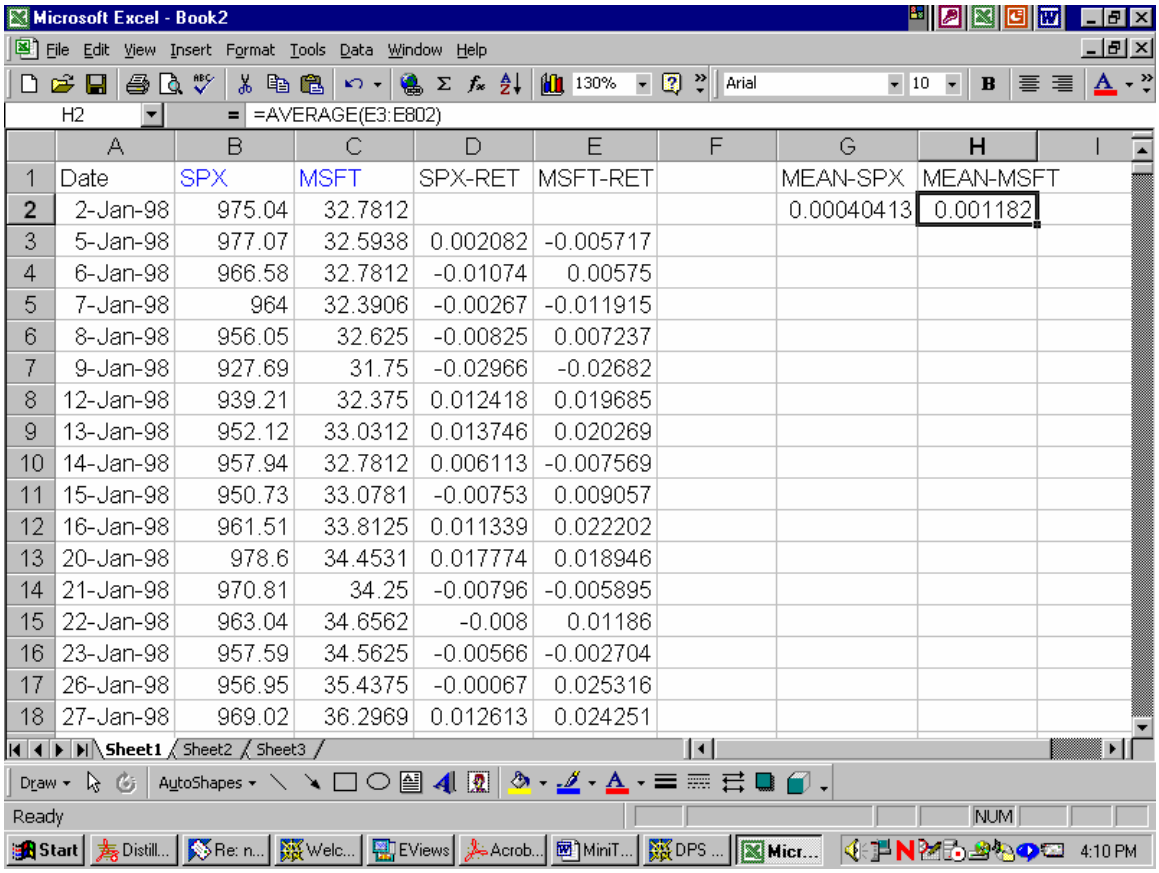
=average(

and then highlight the array of cells containing the values to be averaged, here D3:D802.



	A	B	C	D	E	F	G	H	I
1	Date	SPX	MSFT	SPX-RET	MSFT-RET		MEAN-SPX		
2	2-Jan-98	975.04	32.7812				0.00040413		
3	5-Jan-98	977.07	32.5938	0.002082	-0.005717				
4	6-Jan-98	966.58	32.7812	-0.01074	0.00575				
5	7-Jan-98	964	32.3906	-0.00267	-0.011915				
6	8-Jan-98	956.05	32.625	-0.00825	0.007237				
7	9-Jan-98	927.69	31.75	-0.02966	-0.02682				
8	12-Jan-98	939.21	32.375	0.012418	0.019685				
9	13-Jan-98	952.12	33.0312	0.013746	0.020269				
10	14-Jan-98	957.94	32.7812	0.006113	-0.007569				
11	15-Jan-98	950.73	33.0781	-0.00753	0.009057				
12	16-Jan-98	961.51	33.8125	0.011339	0.022202				
13	20-Jan-98	978.6	34.4531	0.017774	0.018946				
14	21-Jan-98	970.81	34.25	-0.00796	-0.005895				
15	22-Jan-98	963.04	34.6562	-0.008	0.01186				
16	23-Jan-98	957.59	34.5625	-0.00566	-0.002704				
17	26-Jan-98	956.95	35.4375	-0.00067	0.025316				
18	27-Jan-98	969.02	36.2969	0.012613	0.024251				

Again, you can drag the formula to the right to get the mean return on Microsoft.



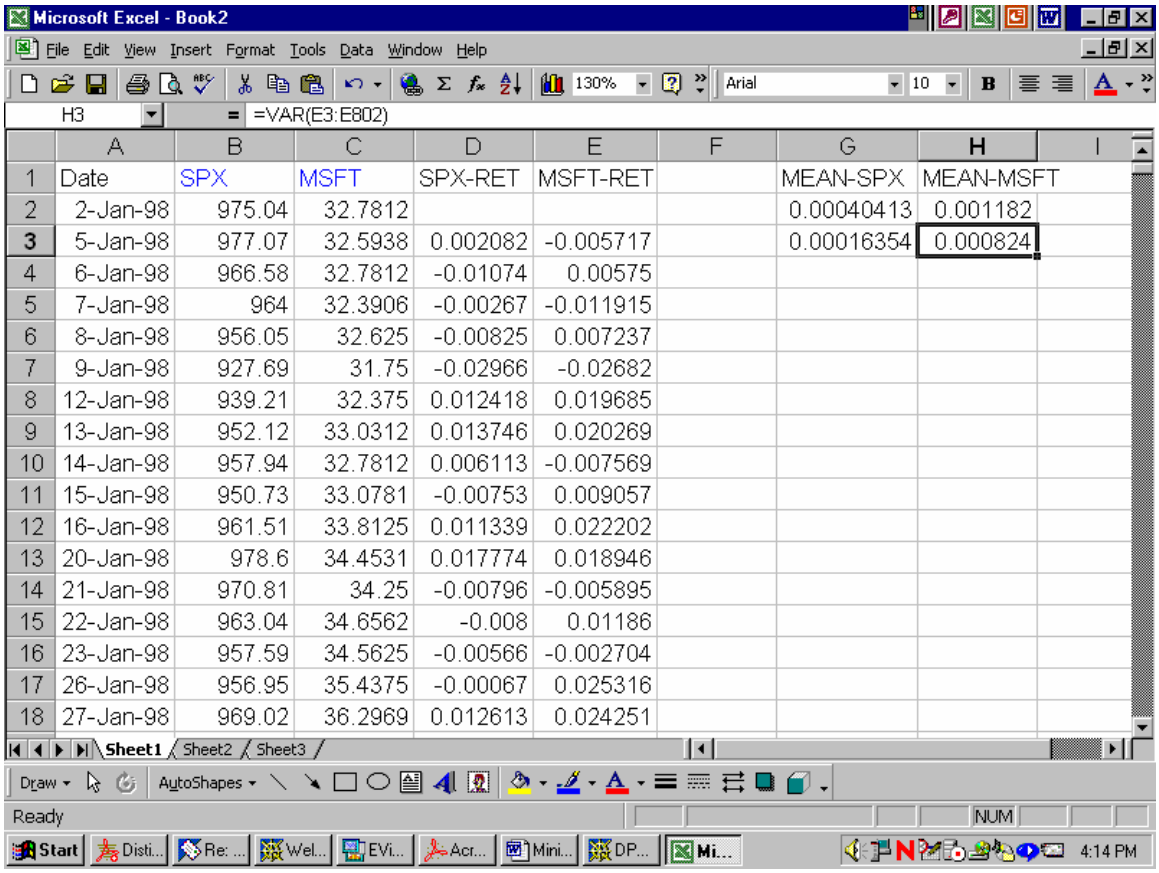
	A	B	C	D	E	F	G	H	I
1	Date	SPX	MSFT	SPX-RET	MSFT-RET		MEAN-SPX	MEAN-MSFT	
2	2-Jan-98	975.04	32.7812				0.00040413	0.001182	
3	5-Jan-98	977.07	32.5938	0.002082	-0.005717				
4	6-Jan-98	966.58	32.7812	-0.01074	0.00575				
5	7-Jan-98	964	32.3906	-0.00267	-0.011915				
6	8-Jan-98	956.05	32.625	-0.00825	0.007237				
7	9-Jan-98	927.69	31.75	-0.02966	-0.02682				
8	12-Jan-98	939.21	32.375	0.012418	0.019685				
9	13-Jan-98	952.12	33.0312	0.013746	0.020269				
10	14-Jan-98	957.94	32.7812	0.006113	-0.007569				
11	15-Jan-98	950.73	33.0781	-0.00753	0.009057				
12	16-Jan-98	961.51	33.8125	0.011339	0.022202				
13	20-Jan-98	978.6	34.4531	0.017774	0.018946				
14	21-Jan-98	970.81	34.25	-0.00796	-0.005895				
15	22-Jan-98	963.04	34.6562	-0.008	0.01186				
16	23-Jan-98	957.59	34.5625	-0.00566	-0.002704				
17	26-Jan-98	956.95	35.4375	-0.00067	0.025316				
18	27-Jan-98	969.02	36.2969	0.012613	0.024251				

The **variance** is the dispersion around the mean:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - m)^2$$

Again, Excel has a built-in function to compute this:

=var(array)



	A	B	C	D	E	F	G	H	I
1	Date	SPX	MSFT	SPX-RET	MSFT-RET		MEAN-SPX	MEAN-MSFT	
2	2-Jan-98	975.04	32.7812				0.00040413	0.001182	
3	5-Jan-98	977.07	32.5938	0.002082	-0.005717		0.00016354	0.000824	
4	6-Jan-98	966.58	32.7812	-0.01074	0.00575				
5	7-Jan-98	964	32.3906	-0.00267	-0.011915				
6	8-Jan-98	956.05	32.625	-0.00825	0.007237				
7	9-Jan-98	927.69	31.75	-0.02966	-0.02682				
8	12-Jan-98	939.21	32.375	0.012418	0.019685				
9	13-Jan-98	952.12	33.0312	0.013746	0.020269				
10	14-Jan-98	957.94	32.7812	0.006113	-0.007569				
11	15-Jan-98	950.73	33.0781	-0.00753	0.009057				
12	16-Jan-98	961.51	33.8125	0.011339	0.022202				
13	20-Jan-98	978.6	34.4531	0.017774	0.018946				
14	21-Jan-98	970.81	34.25	-0.00796	-0.005895				
15	22-Jan-98	963.04	34.6562	-0.008	0.01186				
16	23-Jan-98	957.59	34.5625	-0.00566	-0.002704				
17	26-Jan-98	956.95	35.4375	-0.00067	0.025316				
18	27-Jan-98	969.02	36.2969	0.012613	0.024251				

Therefore, it can be seen that Microsoft has a higher average return than the S&P Index, but it also has more variance than the index.

Remark: For statistical reasons, it is more accurate to estimate variance as

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - m)^2$$

Notice that we are dividing by n-1, not n. This is the formula that Excel uses when you type =var(). If you want to use the first formula (where we divide by n), you should type =varp(). In practice, the difference between var() and varp() is small, especially when n is large. You should use var() in most applications.

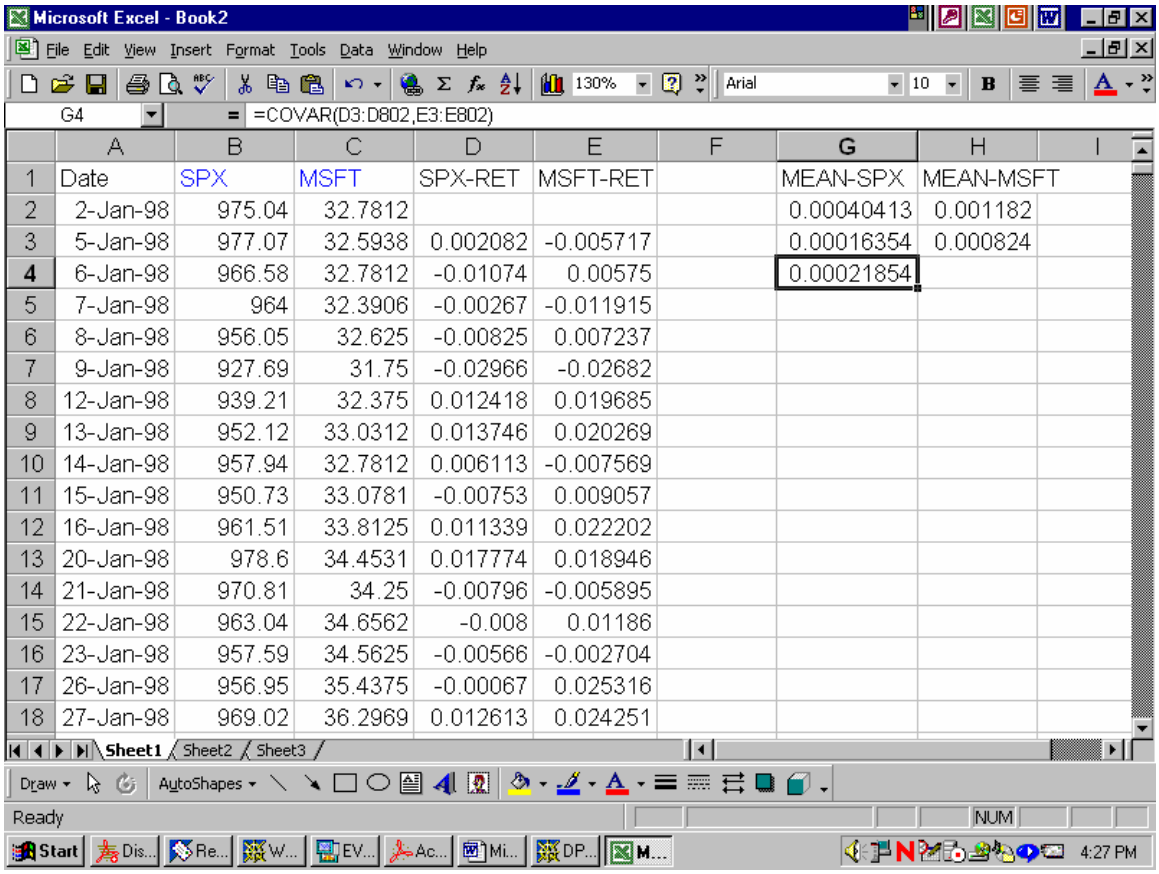
Covariance is defined as

$$\mathrm{cov}(a,b) = \sum_{i=1}^{n} \frac{1}{n}(a_i - m_a)(b_i - m_b)$$

which is easily computed in Excel by using the built in function

=covar(array1, array2)

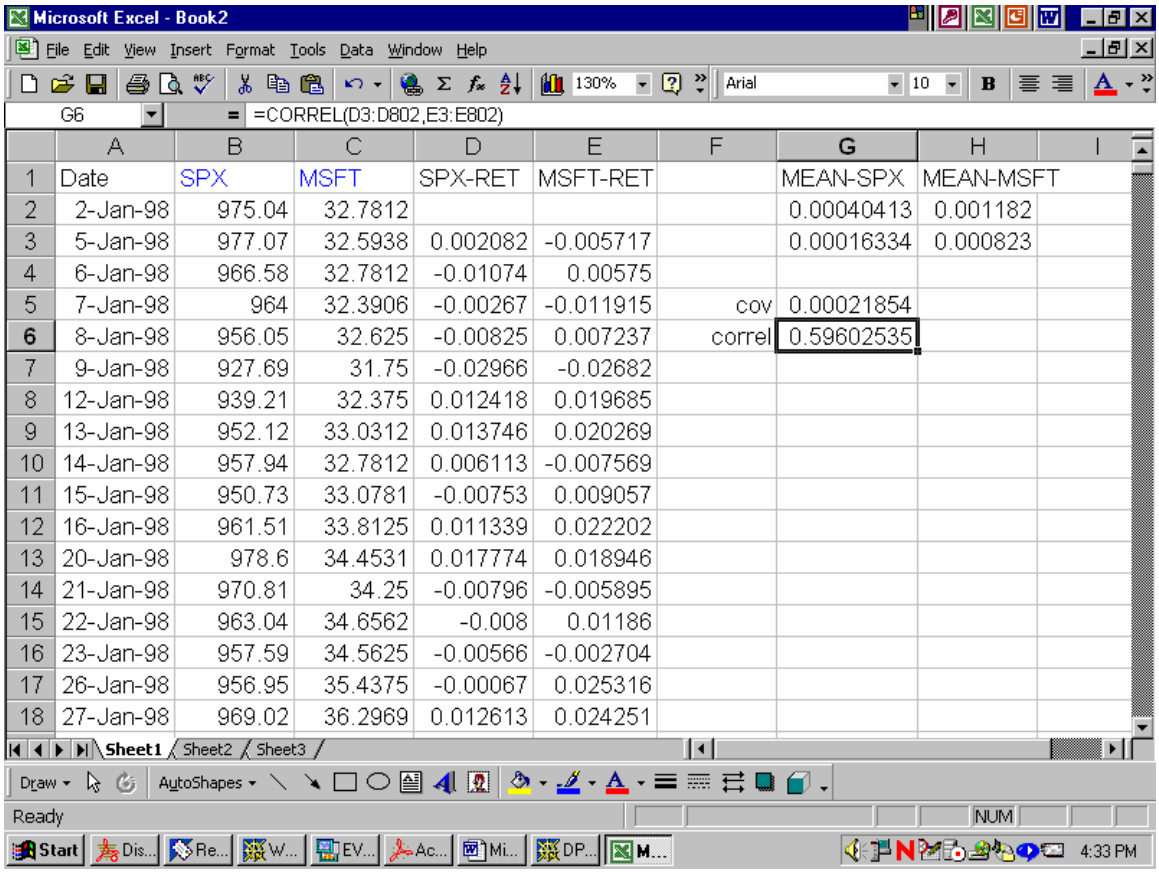
where array1 and array2 must be of the same size.



	A	B	C	D	E	F	G	H	I
1	Date	SPX	MSFT	SPX-RET	MSFT-RET		MEAN-SPX	MEAN-MSFT	
2	2-Jan-98	975.04	32.7812				0.00040413	0.001182	
3	5-Jan-98	977.07	32.5938	0.002082	-0.005717		0.00016354	0.000824	
4	6-Jan-98	966.58	32.7812	-0.01074	0.00575		0.00021854		
5	7-Jan-98	964	32.3906	-0.00267	-0.011915				
6	8-Jan-98	956.05	32.625	-0.00825	0.007237				
7	9-Jan-98	927.69	31.75	-0.02966	-0.02682				
8	12-Jan-98	939.21	32.375	0.012418	0.019685				
9	13-Jan-98	952.12	33.0312	0.013746	0.020269				
10	14-Jan-98	957.94	32.7812	0.006113	-0.007569				
11	15-Jan-98	950.73	33.0781	-0.00753	0.009057				
12	16-Jan-98	961.51	33.8125	0.011339	0.022202				
13	20-Jan-98	978.6	34.4531	0.017774	0.018946				
14	21-Jan-98	970.81	34.25	-0.00796	-0.005895				
15	22-Jan-98	963.04	34.6562	-0.008	0.01186				
16	23-Jan-98	957.59	34.5625	-0.00566	-0.002704				
17	26-Jan-98	956.95	35.4375	-0.00067	0.025316				
18	27-Jan-98	969.02	36.2969	0.012613	0.024251				

To compute the correlation, you can use the built in formula

=correl(array1, array2),



	A	B	C	D	E	F	G	H	I
1	Date	SPX	MSFT	SPX-RET	MSFT-RET		MEAN-SPX	MEAN-MSFT	
2	2-Jan-98	975.04	32.7812				0.00040413	0.001182	
3	5-Jan-98	977.07	32.5938	0.002082	-0.005717		0.00016334	0.000823	
4	6-Jan-98	966.58	32.7812	-0.01074	0.00575				
5	7-Jan-98	964	32.3906	-0.00267	-0.011915	cov	0.00021854		
6	8-Jan-98	956.05	32.625	-0.00825	0.007237	correl	0.59602535		
7	9-Jan-98	927.69	31.75	-0.02966	-0.02682				
8	12-Jan-98	939.21	32.375	0.012418	0.019685				
9	13-Jan-98	952.12	33.0312	0.013746	0.020269				
10	14-Jan-98	957.94	32.7812	0.006113	-0.007569				
11	15-Jan-98	950.73	33.0781	-0.00753	0.009057				
12	16-Jan-98	961.51	33.8125	0.011339	0.022202				
13	20-Jan-98	978.6	34.4531	0.017774	0.018946				
14	21-Jan-98	970.81	34.25	-0.00796	-0.005895				
15	22-Jan-98	963.04	34.6562	-0.008	0.01186				
16	23-Jan-98	957.59	34.5625	-0.00566	-0.002704				
17	26-Jan-98	956.95	35.4375	-0.00067	0.025316				
18	27-Jan-98	969.02	36.2969	0.012613	0.024251				