Text Lecture 3.3 – Coherent measures of risk and back-testing

Dear all, welcome back.

In this class we will discuss one of the main drawbacks of Value-at-Risk, that is to say the fact that the VaR, as a measure of risk, is often not coherent.

But what does that mean?

In general terms, a risk measure is a function that it is used to quantify risk.

It is meant to determine the amount of an asset, say money, to be kept in reserve. The aim of a reserve is to guarantee the presence of capital that can be used as a (partial) cover if the risky event manifests itself, generating a loss. Do you remember what we have said during week 1?

From a mathematical point of view, a measure of risk is a function phi that maps from the linear space of losses to the extended real line.

The space of losses is linear, this simply means that is allows operations such as addition and scalar multiplication

When we associate our measure phi to a loss, what we get is a number (notice that we consider plus and minus infinity as numbers), a number that we can use to determine our reserves.

A measure of risk is said coherent if it is monotone, sub-additive, positive homogeneous and translation invariant.

The mathematical definitions are given on screen.

Monotonicity means that, if Z1 and Z2 are two losses and Z1 is smaller than Z2, then the value of the risk measure in Z2 is greater than the value of the risk measure in Z1.

Sub-additivity means that the risk measure assigned to a combination of losses (a portfolio) is smaller than the sum of the risk measures assigned to the single losses.

Positive homogeneity says that, for a nonnegative scalar "a", phi(aZ)=a*phi(Z).

Finally, for translation invariance we can give two definitions. A mathematical one, in which adding a constant to a loss does not change its risk profile; and a second one, more economical, in which adding a scalar to a loss reduces risk by the same quantity.

If one of these properties is not respected, a measure of risk is not coherent.

Sometimes, we can also assume the risk measure to be normalized, even if this is not strictly necessary for a risk measure to be coherent.

But what are the economic interpretations of these properties?

Monotonicity tells us that it is always possible to have an ordering of losses and of their risk profiles.

Sub-additivity guarantees that our coherent risk measure is in favor of diversification of risk. The risk of a portfolio of loans or securities (and therefore of their potential losses) should be smaller than the simple sum of single risks.

Positive homogeneity simply tells us that risk is proportional: if your portfolio may lose 10 dollars, and you double the amount of money in that portfolio, you may expect a loss of at least 20 dollars.

Finally we have translation invariance. In that case the idea is that, if we add a non risky quantity to our risky elements, risk cannot increase. From a strictly economic point of view it can even decrease! Why?

Imagine you add liquidity to a portfolio. In the short run, we can assume money to be risk free, that is we can ignore inflation. Now, the overall risk of your portfolio decreases, since it already contains some liquidity that can be used to hedge risk.

Let us now consider an exercise, to show that, unfortunately, the VaR is not always coherent.

In particular we show that the VaR is not generally sub-additive, hence it appears to be against diversification.

We have two independent portfolios of bonds. They both have a probability of 0.02 of a loss of 10 million pounds and a probability of 0.98 of a loss of 1 million pounds over a 1-year time window.

The 97.5% VaR is 1 million for each investment.

What is the joint VaR of a portfolio made of the two investments?

Since we are assuming independence, computations are rather simple.

Let's start by writing down the usual table of losses and their probabilities.

In this case it is easy to see that for each single investment, the 97.5% VaR is 1 million.

97.5% is in fact smaller than 98%, the probability of 1 million loss.

Now, let us build another table containing the portfolio that combines the two investments.

The smallest loss we can expect is in the case in which the two independent investments make us lose 1 million each. The total is therefore 2 million pounds. The probability of this event is 0.98 times 0.98, equal to 0.9604.

Then we have a possible joint loss of 11. This can be obtained as 1 million loss from the first investment and 10 from the second, or vice versa. Hence the probability is 0.98 times 0.02 times 2.

Finally, the worst scenario is the one in which we lose 20 million. Here the probability is 0.0004.

As usual we can draw a graph of the empirical cumulative distribution function of losses.

On the x-axis we have losses, while on the y-axis we have the cumulative probabilities.

The 97.5% VaR of the portfolio is obtained by drawing the usual horizontal line passing through 0.975.

The line intersects our ecdf at a point, which corresponds to a loss of 11 million pounds. This is our VaR.

Now notice the following: the joint VaR is 11 million pounds. This quantity is definitely larger than the sum of the two VaRs of the single investments, which is 1+1=2 million pounds.

This simple example shows that the VaR is not sub-additive, hence it is not coherent.

Ok. And what about the expected shortfall?

In that case we can easily show that the ES is always coherent and, in particular, it is always sub-additive.

It is a nice exercise to show that on your own.

In general, we can say that the ES is always coherent, while the VaR is not.

However, there are special cases in which the VaR is coherent, since it shows to be sub-additive. For example, if the loss distribution is a Gaussian distribution, it is not difficult to show that the VaR is coherent. This holds for all elliptical distributions.

Can you think of a simple example showing this?

Ok, let's now change topic and consider back-testing.

Back-testing is a validation procedure often used in risk management.

The idea is simply to verify the performances of the chosen risk measure on historical data.

Consider a 99% C-VaR. What we want to check is how that value would have performed if used in the past.

The idea is to count the number of days in the past in which the actual loss was larger than our 99% VaR.

If these days, called exceptions, are less than 1% (1-alpha) then our back-testing is successful.

If exceptions are say 5%, then we are probably underestimating the actual VaR.

Naturally it may also happen that we overestimate VaR!

Standard back-testing is based on Bernoulli trials generating binomial random variables.

If our VaR alpha is accurate, the probability p of observing an exception is 1-alpha.

Now, suppose we look at a total of n days and that we observe m<n exceptions.

What we want to compare is the ratio m/n and the probability p.

To perform this test we use the binomial distribution.

The binomial distribution tells us that the probability of observing more than m "exceptions" in n "days" is the one you see on screen.

The idea is that, once we have chosen a significance level for the test, say 5%, we look for the right tail probability of our binomial. If this probability is smaller than the chosen significance level, we reject the null hypothesis that our VaR is ok, thus rejecting the VaR.

If the probability is greater than the significance level, the VaR appears to be ok.

Notice that this is true if we test about the underestimation of VaR. In the case of overestimation, we are interested in the left tail, but the philosophy is the same.

An exercise will make things easier.

We want to back-test our VaR using 900 days of data. Alpha is 0.99 and we observe 12 exceptions.

According to our VaR we expect only 9 exceptions (1% of 900).

Should we reject our VaR?

What happens with 20 exceptions?

We will solve this exercise using R. So…let's go.

In R it is very easy to compute the probability we are interested in, simply by using the pbinom() function.

Since we are interested in a right tail probability, given that we are interested in seeing whether our VaR is underestimated, we can simply type 1-pbinom(11,900,0.01,TRUE).

900 is the number of days, 0.01 is the 1-alpha level, and TRUE is an option we use to tell R that we are interested in the cumulative distribution function of a binomial.

But why are we typing 11 and not 12? Simply because the binomial is a discrete distribution, hence the probability of observing 12 or more exceptions is equal to 1 minus the probability of observing 11 or less exceptions, which is what pbinom(11,900,0.01,TRUE) computes. I refer you to the R help, for more details about the use of the function.

Let's say that we choose 5% as significance level for our test (but it could also be 2% or 1%, or what you prefer). Since 0.1960 is greater than 0.05, we cannot reject the null hypothesis, and this means that our VaR is correct.

In case of 20 exceptions, on the contrary, we obtain 0.00099. This quantity is smaller than 0.05, hence we reject the null hypothesis: our VaR is underestimating risk.

Notice that the results hold true also if we choose a 1% significance level.

As for all tests, the choice of the significance level is very important.

In the case of overestimation, we are no longer interested in the right tail, but in the left one, hence instead of 1-pbinom, we just consider pbinom (the cumulative distribution function).

In other words, we are interested in the probability of observing 2 or less exceptions (4 in the second case).

You will just need a little practice, but the methodology is simple.

If you have problems, do not forget to use the course discussion forum.

Ok, goodbye.