

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](https://ocw.mit.edu).

**PROFESSOR:** Our guest speaker today from Morgan Stanley, Ivan Masyukov. Dr. Ivan Masyukov.

**IVAN MASYUKOV:** Hello. One, two, three. Can you hear me?

**PROFESSOR:** And the microphone will just be recording you, but it doesn't broadcast you.

**IVAN MASYUKOV:** Ah. Understood. All right. So I'm Ivan Masyukov. I work in Morgan Stanley. And my background is applied physics and mathematics from Moscow Institutes of Physics and Technology.

And today, the topic of the lecture is regularized pricing and risk models. So we will talk about typical pricing risk models for interest rate products, and the important aspect of adding some additional constraints, which means, like, adding some regularizers to the model.

So we will start from bonds, which is probably the most simple interest rate product on the market. Then we will discuss swaps. We will build a yield curve. And we will see how yield curve models can be improved to satisfy needs of actual trader.

And at the end, we'll look at the very nice example of ill-posed problem of calibrating the two-dimensional volatility surface necessary for volatility model-- Monte Carlo assimilation. And we will see how that problem can be solved. During the lecture, if you have any questions, please interrupt, OK?

So what is bond? Bond is a security which is issued if someone like a borrower needs money. And it promises to pay some certain fixed amount of certain cash flows in the future, and request for some money up front for this. So typical bonds basically include same periodic payment-- let's say like every half year or every year until maturity, where at maturity the face is paid, like the biggest sum of money. And

again, during the beginning, the investor is asked to pay some up front.

There are also zero-coupon bonds, which don't pay anything until the maturity. And there are very interesting perpetual bonds where basically you pay some money up front, and then you pay it back like infinitely-- which sounds like a good deal, but we will know how to price it right.

So those are some diagrams. So the first one is the standard fixed-rate bond, where small green arrows represent a periodic payment. And there is a face value added on top of a periodic payment at the maturity of the bond.

So this is a typical cash flow diagram used for analysis, OK? And so arrows up represent something that-- and it's green right? That is good for us. So it's something that we receive. And a red arrow facing down represents something that you have to pay. Right?

So a zero coupon bond is something, as I said before, is something that you pay up front, and you get back a fixed amount of money in the future. What's interesting about this graph-- you can see that the green arrow has a bigger amplitude than the red one, which means that you kind of, every time you put like \$100 now, right, you kind of expect that in return you get more in the future. Because if you don't get more in the future, you just don't get this money, don't put this money. You just keep it in the pocket. So as a result, you get the concept of time value of money. So tomorrow, \$100 always will be more than just \$100.

And also, if you look at the graph of the fixed-rate coupon bond, and use all of the cash flows here, it looks like you get more than this red one. But again, there is, as further in the future the cash flow is, the more the kind of depreciation. And we call this depreciation a discount factor, OK? So basically the more in the future the cash flow is, the smaller the discount factor. And so for today the discount factor will be 1, for tomorrow it will be like 0.999, and so forth. And in 30 years, let's say, it will probably be like 0.1, depending on current rates in the market.

So let's see how we can price the bond-- or not necessarily price, but compute a fair

value of future cash flows. So our fair value of computed cash flows can be found if we have discount factors. So every discount factor at every cash flow in the future I- - which in this particular case will be a coupon times the face value-- should be multiplied by the discount factor. And then we also add a face value discount with a discount factor at the maturity of the instrument.

So the way this product trades in the market is that people buy and sell bonds paying P, right? So it's very important to understand that for bonds, it's not something that we have cash flows which we kind of need to price. It's actually the price is already known. So it's very liquid. It's the result of activity in the market, meaning that there is very little uncertainty about the price. So this P is known.

And as with all cash flows, it's something that's written in the contract, right? So it's something, we have fixed cash flows in the future. So it's always a question about what kind of model is useful for the discount factors. So we need a model for discounting.

Any questions so far?

So one of the simplest models is to use just one parameter to kind of cover all the discounting. And the discount factor can be represented as  $e^{-y \cdot t}$ , where y is some kind of-- it's called yield to maturity.

Well, the reason why it's exponential, it's natural right? So if you have a 0.999 discount factor for today, and then we kind of say, OK, it's the same discount for tomorrow, we will have the same discounting for every other day. So we have to multiply them. As a result, the total discounting will be an exponential.

So if our discount factors are like this, then our price basically can be represented as a linear combination of future cash flows, right? At this point, by the way, we kind of merge together the final coupon with the face value, and we'll just kind of be talking about the coupons only, about cash flows only. And so that's the formula for the bond price, is this.

So basically, what's known on the market is P, right, which is a price that's-- that

instrument is straighted. We also have defined cash flows in the future. So we can solve for yield.

So essentially, if we know the bond price, we can find the bond yield, OK? And if we know the bond yield, we can find the bond price, OK? So typically, bonds are traded in terms of its price. But some bonds are traded in terms of yield. But again, this is like one-to-one. You can always go back and forth.

What's important is what has economic value, right, is the bond price, OK, and the future cash amounts of cash flows. And when you talk about yield, it's not something that's straighted. It's actually one of the ways to align future cash flows with the bond price. And that way assumes that we have, again, constant discounting for all time points in the future. And we will see that it may not be optimal case.

So what's also important when we're talking about instrument price is to have the model of how that price changes if the market changes. So here, we're talking about sensitivity of the bond price to yield. And what is typically done is basically to normalize by the bond price itself. And then its called bond duration.

So the nice thing about normalizing is that the duration of that bond that you have in your portfolio doesn't really depend on how many bonds you have, right? So it's basically more like property of the bond itself, rather than how many bonds you have in your portfolio.

So if you take the previous formula and take the derivative with respect to  $I$ , we get the following formula for duration. And we know what the price is, right? And we can rewrite this formula this way, which you see it's a sum of  $t$  times some weights and divided by the sum of the weights. So it's essentially a weighted sum of time, OK? And those pieces of time, those moments of time is more important, as I mean the weights are proportional to present values of future cash flows.

So that's why bond duration has a very nice kind of intuitive sense. As a result of that, and yeah, I forgot to mention one thing. So the duration is always a negative, right? So we have a sign here, because if the bond price goes up, this means that

the yields goes down, OK? And if the yield goes up, price goes down.

And the explanation's very simple. So yield is kind of the same thing as interest rate on the market. So if those rates go up, this means that there will be more discounting in the future cash flows, they will be less valuable to me. So I'll be less willing to pay for those cash flows, OK? So it's kind of fundamental that relationship has a negative sign.

So in case of a zero-coupon bond, we only have one cash flow in the future. So there is just one weight, and that weight is totally assigned to that last cash flow. So duration of zero-coupon bond equals to maturity. Duration of regular coupon bond depends, but it's always less than maturity, just because we'll have a weighted sum formula here.

So essentially, that model for the bond duration kind of assumes that all rates-- so we have just one yield number for everything, so all rates go like in a parallel way, which was OK before the crisis, right, where kind of rates today are kind of similar to the rates expected in the future. But it's no longer the case.

So the rates now, they're higher than like one year ago, but they're still much lower than expected in the future. So we expect that the rates will go very high. So the curve is very steep at the moment. So that model of just one number for everything might not be adequate. And we'll see how we can improve this situation.

So it's worth mentioning the secondary, which if we already spoke about the price, first derivative of the price with respect to yield, and a second. So for small changes in the yield, you can assume that it's linear, so it's OK to use just the first derivative. So second derivative will be necessary for larger movements of the market.

Like as an example, if you're a trader, right, and the bond trades, right? So we call it a cash product. Means that you actually don't need any model to price it. You already have that price, OK? But if you try to explain like why you might have lost money today, right, and that always the trader always does that at the end of the day. And we always use first derivatives.

And we try to explain it, but there is also unexplained, OK? And that unexplained can be quite high on this with large movements. So if you have like in term analytics for the bond convexity, that helps you to include the second derivative, and therefore make the second derivative smaller-- sorry-- the unexplained smaller.

So let's now talk about interest rate swaps. So bond cash flow is basically a stream of fixed cash flow, which means that for certain dates it's just guaranteed that you will be getting \$100 with certain periodicity. A swap means you exchange fixed payments with respect to some floating. And floating means that the amount of money that you'll be getting or paying, OK, receiving or paying, will depend on some market observable.

So for interest rate swap, it will be typically-- and let's focus on the USD market, it will be a three month LIBOR rate. That rate is published daily, OK? And it's like if you need to go to the bank and get a three month CD with the money for the three months, that rate is already known. It's actually called LIBOR, because it's kind of between banks, and it's set at 11 AM time, London time.

So as a result, we already know how to price cash flows in the future. So present value of the fixed stream of payments, as we know, will be like this. And there is a floating rate of cash flow as well.

And the nice thing about the swap is that when you enter the swap, you don't pay any money, right? It's because you just kind of enter the agreement, rather than when you buy or sell a bond, there is some exchange of money. For swaps, swaps are designed the way such that so when you make this agreement, it's a certain moment of time, the fixed rate of the swap is picked in such a way that the present value of fixed minus the floating cash flows will be net to zero.

So you can see, I mean, if we rewrite those equations, OK, we can see that the swap rate-- which is the most important quantity of the swap, and something that traders are basically are most concerned, right? So you first need to define what the swap is. And for USD, you are saying probably like 10 years swap, OK. And this is the rate. So the trader continuously kind of quotes bid of levels of the swap rate. So

one is talking about PVs and stuff like that. So it's always the swap rate. So that the swap rate is weighted sum of forward trades.

And it has a very nice intuitive explanation. So you have some stream of floating cash flows-- variable cash flows-- which at the moment, like, will probably be low now, but will be high in 10 years, will be much higher in 30 years. So the swap rate for this kind of environment will be kind of an average, right? And again, those weights depend on the discounting factors.

So later, we will see that because we're talking about bond, that having a fixed cash flows in the futures, and a swap that fixed in exchange of floating, swap can be hedged with bond, "hedged" meaning that-- you know what the term "hedged" means, no?

Hedging means that if you have just, let's say, a swap, right? So if market changes, right, you can again lose money. So a typical task for the money-maker, trader, is to kind of offset that risk with something. Ideally, you sold one swap, you bought another swap the same way with a different rate. So you kind of locked in your profit. But you remain with a zero risk.

So let's try to construct a yield curve. Why do we need the yield curve? So when we have, let's say, a series of swap with different maturities, right, all those swaps will start today, and usually, swap will have quarterly payment for the floating leg, and six month payments for the fixed leg, and you'll have different maturities.

But if you try to kind of get discount factors from that information, you will see that you can get those discount factors only for certain dates, OK? But the typical situation is that given on some liquid market instruments, you want to price your entire portfolio, which has continuous spectrum of cash flows from now to 30 years, 40 years. And also, for typical swap portfolio that I personally deal with on a daily basis contains hundreds of thousands of swaps. Every swap has many cash flows.

So you need something that can, based on discrete information of reliable liquid instruments on the market, draw the line. Can basically construct the curve. Which

means that you can, so that you are able to get discount factors for any potential day in the future, or you can compute forward trade for every date in the future.

So the first step to construct a yield curve is to select input instruments for calibration. So you have a set of instruments, and a new set of input cause. Then you also need to kind of decide what kind of properties of that line will be. So you can say, OK, first of all, you need to decide what quantity will be interpolated. It could be daily discount factors, or daily forward trades, or maybe three-month forward trades.

Then you select the spline [INAUDIBLE]. So I'm not sure if you're familiar with the splines. Probably you heard about cubic spline, right? So there are different types of splines, and some of them are better and some of them are worse for different situation. And you also need to decide like what that will be the node points for the spline itself.

OK. And then, as a final step, so you have some mathematical quantity, which is mathematical object where you know what the line is, and you have control points. And you need to adjust your control points such that when you reprice your instruments, those instruments are repriced exactly to the same course that you find on the market. You have a question.

**AUDIENCE:** Is that spline, again, is it just like a--

**IVAN MASYUKOV:** All right. So let me. All right. So this is a picture of the cubic spline. So spline is a way to draw a smooth curve. This is an example of the cubic spline.

So you start to define your node points. Your node points in this case are 1, 10, 20, 40, 80, 160, and 240, right? And then for every one of those intervals, OK, the functional form of the shape of this curve is a cubic polynomial, OK?

Well, if you just do cubic polynomial for every interval without kind of putting additional constraints, you can have all kinds of boundary effects, like jumps, kinks, and other things, because we want our cubic curve-- cubic spline-- to be meaningful, right? So we want to maintain to preserve maximum number of



derivatives for every node point.

So we're not going to check. But believe me, this curve, it is a cubic polynomial for every one of those interval. And also, we have two continuous derivative at every node point, because for the  $n$  degree of the spline, you always have like  $n - 1$ . You can have  $n - 1$ .

So the same thing, a spline can be represented in terms of B-splines. B-spline is a new type of spline. It's just as a representation which is more intuitive, I should say. So all universe of the curves with those node points, max with two continuous derivatives, can be represented as a linear combination of those basics functions. So B-spline, I mean, if you're interested, you should probably, we're not going to discuss it in details, but it's nice separate kind of topic about how to build those B-splines.

But essentially, what's nice about those B-splines-- and "B," as you probably already understood, B is basis, right? So you have like basis functions. So those functions look like bell shapes. They are non-zero on some sub-interval. On every interval, it will be a cubic polynomial. Everyone will have always two continuous derivatives. As a result, in any linear combination of those-- which the first curve is-- will also have that property.

OK. So now we, yeah, so the calibrate means that we basically have some solver to make sure that our swaps with the rates for those maturities actually re-price at par. At par means that the PV is zero.

This is a typical example of the yield curve instruments. And IRS stands for "interest rate swap," and we have maturities from one year to 30 years. And the quotes are of 0.33% up to 2.67%. So you can see that that's actually, that's one from my one-year-old presentation. Rates are quite high these days.

So this is an example of the yield curve graph. So again, so those are the rates from 0.3 to 3.5.

And the shape of the curve is not flat at all, right? So it's actually pretty steep. So for

the first five years, it's very steep. Then it reaches the plateau. And then there is some feature there, probably because of some behavior in the 20-year region.

So three-month forward rate is the LIBOR rate. LIBOR is the rate for the three month. It's mostly kind of common. And the reason why is because the standard interest rate, usually this swap has a three-month frequency for payment on the floating leg. So if you're talking about floating rates, is always three months. And it's always LIBOR.

So because we've already built the curve, now let's see how we can improve the situation with a bond. So we have the curve, so we have the discount factors, right? And we see that those discount factors cannot be obtained on the assumption that you have just one parameter yield for everything, because the curve we know is not flat.

So if we just try to price it using those discount factors, try to get a fair price, we probably won't match the market observables. So we need some extra term. And again, here we can use it in a similar form as we did it for the yield. But right now it's going to be a small correction to the yield curve, rather than kind of really rough assumption about that the curve is flat, OK?

So typically, if the curve magnitude is, let's say, 3%. OK? So the spread is probably 100 times lower. So having a nice correction is always better, right?

And another nice feature is that of this approach for the bond, like if we already build our yield curve model, and we know sensitivities of our portfolio to inputs of the curves, which then transition into like differences in discount factors, we can easily apply that to the bond. We can first find what this spread parameter is to solve for  $s$  knowing  $P$ , which is very liquid market tradable. And then we can kind of use consistent model for the bonds and the swaps in our portfolio. Any questions?

**AUDIENCE:** Yes. So what does the bond rate tell us about the bond?

**IVAN MASYUKOV:** That's a very good question. So it might tell us something like bond liquidity, for

example. Like if it's not liquid, or there is some-- so it may be related to the bond itself.

And sometimes we kind of think that the bond is riskless, which means that-- especially if it's issued by US government, which if we can assume that those cash flows in the futures are guaranteed, right, then I basically will be willing to bring a certain amount of money and discount factors, right? But if you tell me that you will pay me that in the future, I won't be so certain, right? So I'll need to add some kind of credit spread to that-- we call it credit spreads-- as a result. It's the credit spread will kind of propagate to the spread number.

On the other hand, if the bond is really US government-issued, and is considered to be guaranteed, then it may be a feature of the swap, OK? Where just because of some liquidity situations in swap market-- like all of a sudden, let's say, all option traders on the street needed this 10-year swap, OK? Because they kind of need to hedge certain very popular products-- volatility products-- they'll start to buy it, that spread will change. But what's even more interesting, that spread is tradable by itself, OK? So you can go to the market and you trade the spread.

Moreover, let's look, like, 10-year situation. So you have 10-year bond on the market. You have tradable swap, and you have tradable spread. So the question is which one is the most liquid? What do you think?

The most liquid is the bond, of course. It has much more liquidity. Surprisingly, the second one is the spread between the 10-year swap, and the bond is traded in the market. So there's more transaction on the spread rather compared to the swap.

As a result, when we built our curves, we're not taking like 10-year swap from the market. OK. We actually take the yield and the spread. And that's how we define the most kind of reliable level of the swap. Of course, we could have just take whatever we observe for the 10-year swap, but it could be off. And also, if you observe, there will be more a bit of a spread as well.

So as an example, let's try to shift one of the inputs of the curve by one basis point. And that will result in this kind of deviation of forward trades, which will be combination of basis plans. But what's interesting first of all, it's kind of complicated [INAUDIBLE] behavior. The reason why is because you are saying that nothing changed before the nine year, like nothing changed after ninth year, but just point in between. So in order to calibrate to that kind of weird condition, right, you need to have a ripple here.

But what also is more important is that by shifting one year basis point by one basis point, that the amplitude of shifts in the curve reaches 14 basis points. So not sure if you're familiar, but it's an ill-posed problem right? So small changes in your inputs can cause large variations in your outputs.

This is a very important slide. So the first column is, again, we saw those are our instruments, quotes, and this is the risk of the portfolio. That's something that a trader needs no matter what. It basically shows you what happens on the market if different-- what will be the change in your portfolio if the market changes. So the meaning of the number-- for example, for the five year-- is that if five year rate moves up by one basis point, we'll lose minus 1,700K.

We also marked here yellow points that are more liquid than the others. So now a typical situation is that you need to hedge your portfolio, right? So you need to liquidate your risk, I'm basically saying, OK? Given the model that we have, I want its value to be insensitive to any movements on the market. So for that purpose, what you can do, you can go and you can buy as many one-year swaps plus 200, as many two-year swaps which would be the risk of minus 1.3, and so forth, right?

Then that always cost you money right? And that money is kind of proportional to bid offer of the particular instruments. And that bid offer is smaller for liquid instruments and larger for less liquid instruments. So if you multiply by the diff-- we can see that if we want to hedge our risk, it's going to be quite expensive. It will cost us 3.6 million dollars. Any questions so far?

So traders never hedge every bucket in the risk. Bucket means every line here. So you always see some numbers, but if you try to make every number here zero, which means that if you trade seven here, you also could try to go to the market, find the offsetting seven here, you'll have to pay too much and you won't be profitable. So what the traders do if someone ask for the seven year, they make this transaction, but they go, then hedge it from the more liquid points, which is less expensive to buy.

So we need a better model for hedging. And a general formulation of the model is presented here. So we have portfolio risk which is a just the vector here, right? And we have hedging. Portfolio risk is basically if you have candidates of instruments that you can use for portfolio hedging, again, the risk will be represented in this format in terms of sensitivities to swap rates.

And we have weights of that hedging portfolio that we need to find, obviously. So you have this hedging portfolio. You multiply  $H$  by  $X$ , you get risk of this hedging portfolio. You add it to the risk of your portfolio.

And then, what we need to minimize, you don't need to minimize everything. But you need to if they give you, OK? What can happen on the market? What are the typical movements of the market? And so essentially you cannot define your market scenarios, which can be found in a different way.

So one of the ways to approach that problem is to use principal component analysis. I know you already are familiar with SVD.

So if a  $D$  is date of market movements in matrix-- then any matrix can be composed using SVD. And we can then look at this spectrum of this decomposition looking at those eigenvalues, and just pick the ones that look high enough for us, and just skip that number.

And let's, for example, we find that we really investigated this market, and we found that there are just five components that drive the market, and the rest is just so little that it's meaningless, right? On every day, and we are certain that it's just five

components, five modes of market moments. Then, if we have a curve that consists of 20 points, we don't need to hedge every swap with its corresponding maturity. We can just pick five swaps that are liquid enough and cheap enough for us to hedge, and just use them.

So let's look now at those typical graphs of those principal components. X-axis is the swap maturity in years, and Y is some kind of relative. Let's think of that as basis points. So blue line is the first component which is the prevalent.

And it kind of, you can see that swap rates, they're basically flattish after 10 years, but the first component is pretty steep. And what it says as well is that the main behavior of the market is that rates now do not move, but they will move in the future. And that's basically because Fed is in a hold, right? So they kind of stimulate the market in a way such that the rate remains the same until sometime in the future.

Mode number two is a kind of like tilting situation. Mode number three is more complex. And we'll have several other modes here as well.

So now, following our previous general approach to the problem, we formulated here as-- so we have PCA factors here in  $P$ . And now, because the number of factors that we selected is the number of hedging instruments-- we no longer need to minimize. We can always feed, which you can always achieve like perfect minimization. We can always achieve that zero. So that's why we formulate it as zero. So solving that problem here.

Yeah. And the hedging matrix, this is an example of hedging matrix. So what that matrix says is that if I take one year swap and put in my portfolio, empty portfolio, and then they apply my model, I'll have just sensitivity to that particular swap. Which kind of makes sense, because since you use the same instruments to calibrate your  $U$  curve, right, then it should be sensitive to itself only. That's why that matrix is just once for itself, and zeroes otherwise.

So then, as a result, we get this matrix. So same portfolio that we had before. This

is our PCA matrix that translates our risk into those few numbers, right? And because we know it translates our risk-- low risk, in terms of many curve inputs-- into just five most liquid ones, which is 1, 2, 5, 10, and 30. As a result, our translated risk, which tells us what we need to do to hedge our portfolio, is just those numbers.

And now, if we take a bit of a charge, 0.1 basis points for those, and multiply, we get numbers which are orders of magnitude smaller than we got before, right? So we probably get something like there were 400. It's not 4 million-- 3.6 million anymore. That's exactly what traders do. And different traders have different opinions of what dynamics of the market is. But they always have some model.

So disadvantages. So PCA model is something that just formally attuned to historical data. I always say that if you take kind of scramble your swap maturities in your model, and you do your computations, and you kind of unscramble them, you get exactly the same result. Which means that in PCA model, you don't put any constraints on that. Two year is very close to one year, and two year is between one year and five year.

So PCA model of hedging coefficients of that matrix is not very stable-- especially for recent modes in the market. Also, because SVDs kind of in the list is the least squares approximation, it's very sensitive to outliers. So there is just one event on the market that kind of one day happens, something like rates went up, and then it went down significantly, it may have necessarily high influence on the outputs.

And if those coefficients change daily, right, then again, it may be too costly. And quite often we just overfitting to historical data. So we're saying, OK, what can I do. I just take historical data, and I prove that my model works, would have worked for the last three years, or the last three months, but that doesn't mean that it will work for the next three months.

If we kind of try to put some additional constraints or additional thoughts about what this behavior should be, this may improve situation. So PCA interpretation is that risk metrics is a linear combination of principal components producing a shift on one

hedging instrument at a time.

Now the question is, let's forget about historical, OK? Is there any other approach we know historically is noisy, and it's kind of first step if you want to do the model. But can we do something better? And the answer is yes.

So we can say that we have our yield curve in terms of forward rates. And typically, when we build this curve, we observe that it is smooth. It's smooth not only because we use smooth splines, but also because if there is no certainty about some event in 10 years from now, there is no reason to kind of expect there will be spike or some non-smooth feature in the formal trade space.

So what we can do is that we can try to minimize those equations where Jacobian is a matrix translating shifts of yield curve inputs into movements of forward trades. So essentially, we will try to penalize non-smoothness.

And the solution will be like this, with some kind of-- so we'll be adding a penalty, OK. And penalty will be a small regularization parameter. So this is, as an example, that's what we get.

Again, here in that model. You can view this matrix as if one year rate moves, what it basically-- so your drivers are 1, 2, 5, 10s, and 30s. So that's your drivers. Knowing the moments of your drivers, what would be the response to your swap rates? And you know that it always be one to itself, right, as you see here. And in between, it will be kind of a smooth functions.

So let's take this moment a broader view at what the pricing model does. And we have a pricing engine, essentially. It's a way, if you have all model parameters-- including curves, volatility, the surface, everything, right? And in order for those parameters to be consistent with the benchmark prices, you need some calibration engine which measures market observables to the ones that's being repriced by the model output of the pricing engine.

And once you make sure that benchmark prices of your model equal or are kind of



close enough to benchmark prices observed in the market, you calibrated the model, then you can essentially price your portfolio, and get values and risk.

So let's look at one of the nice examples of how that pricing engine and pricing and calibration process works. We'll look at HJM model, which is used to price volatility products.

So we're not going to go into too many details about this, but this is a question of evolution of forward rates that we need for simulation from Monte Carlo assimilation. What we're saying here is this change of the forward rates-- because forward rates is the quantity that is being assimilated-- has some drift, OK? Because  $dt$  is time. And also, it all depends on the forward rates to the power of  $\beta$ , right? So if it's log normal model,  $\beta$  will be one. If it's normal model,  $\beta$  will be zero. But in general, it's different.

Then we have volatility surface, right, which kind of gives you what the number of volatility to use for this calendar and forward time. And we have correlation and factor structure which we're not going to talk about here. And this is Brownian motions.

So we're not going to go any more complex like this. We'll just start looking at nice, two-dimensional surfaces here, and see what are the problems of calibrating the volatility surface.

Just to give you a diagram of when we look at the surface, what different elements of that surface mean. It's a triangular surface. You have a calendar time, right, and you have a forward time.

So the assimilation starts at the first vertical line. So you have forward rates here as calibrated from the curve as of today. So those are square boxes here, square elements.

So you need to kind of transition from the first line to the second step using Monte Carlo assimilation. And that's when, for every arrow here, you need the volatility

number. Then, once you did your Monte Carlo assimilation for the second one, you need ones for the third one. And again you need data-- which volatility to use, OK? So essentially, the surface that we'll be looking at on next slides is essentially representing the numbers necessary for this transition's volatility for every arrow.

So to explain, there are different areas here. Like, for example, those will be if one step is one here, OK, we're talking about this, it will be the forward rates that will be observed in two years. That one will be observed as of now for like one year from now. But again, in two years.

And so those rates are essential to compute the swap, forward swap rate. And if we do our Monte Carlo assimilation, that's essential information that we need to compute the price for the option on the swap, which we're not going to discuss here. But just an example, it shows that for different instruments observed in the market you have quite overlapping areas of sensitivity.

So this is a typical example of the volatility surface where this is calendar time, this is forward time. And it has spikes for certain regions. But in general, it's smooth.

So why this problem is challenging. If we try to compute the triangle matrix, which has dimension of 240 by 240-- the reason why is 240 is because every element is for three months, OK?

But we need up to 60 years of data, which means that it's 60 by 4, which is the number of quarters, is 240 by 240. If you just need triangle elements, it's 200K elements. So if you try to calibrate everything at the same time, and you formally try to solve your problem, you kind of needed to store, at least to build a matrix of 28K by 28K. And we just don't have memory for this.

And we also have very small number of calibration instruments only in terms of swaptions or caps, which are typical volatility products. We just have a relatively small number. So it's an undetermined problem. We also, as we saw on the previous example, areas of sensitivity of different instruments overlap. And it's an ill-posed inverse problems which produces unstable solutions.

And no matter what we do, right, the resulting surface should be nice, right? Should look nice, right? Because if it has spikes in some points in the future, then we either have an economic reason for this, or we claim that this is something that's not realistic.

So this is how we approach the problem. So the first step, we represent our volatility surface. And here, even though volatility surface is two-dimensional we just kind of assign a number for every of those elements, OK, and then represent the surface as a vector, OK? Whereas saying that the new surface  $v$  will be some initial state plus a linear combination of basis functions. And basis functions should correspond to some reasonable functions, OK?

But the nice feature is that number of basis functions will be much smaller than the number of elements that we need to calibrate. But after, we will be very formal here. And we'll try to use same number of basis functions as we have our input instruments. So in case we had 50 input instruments, we select basis functions also the number 50.

So we will use typical Newton-Raphson approach here. We will compute sensitivities of input of all instruments to perturbations of a volatility surface, OK? We'll build this Jacobian matrix.

And then, if we made the reasonable assumptions about what those basis functions are, then we can invert our square Jacobian. And again, the reason why a square is because we selected same number of basis functions as the number of input instruments. It's actually quite common approach, but it's very often is wrong approach. It produces unstable results. And we will see why.

So we converge to exact solution, but now the volatility surface looks like this. It looks less like a volatility surface, but more like Manhattan skyline. So you have a Hudson River here, and you have some buildings right? Obviously, even though it calibrates exactly, right?

And you could go and price your portfolio, but probably prices for instruments in the

portfolio that are not input instruments for calibration would be meaningless. Because the reason why we need this surface to be smooth is because for similar instruments for similar products in your portfolio you kind of expect similar prices, right? So if your volatility jumps, that's something that just contradicts with this assumption.

So now how can we improve the situation? So we can try to use our basis functions which were selected in terms of piece by constant shift of different areas. We can use a smooth version of those plans. But again, the result looks better, but still is not good enough.

And just to demonstrate that this is an ill-posed problem-- an ill-posed problem is something that small changes of your inputs results in insane changes in your output. And this is a typical example.

So keep all the instruments the same, OK? We just change by 1%-- which is not a big number-- of the five year by 10 swaption, results in the quite large change of the volatility surface. But look also at the shape right? So it's really kind of you look at one building with an antenna, and another building, right? So it's very unreasonable change of the volatility surface.

So we can use ill-posedness to our advantage. So basically, at this point we say, well, it's not a requirement to calibrate exactly, just because every instruments that is an input of collaboration actually has some tolerance. So even there is no point to calibrate it exactly.

So because we know that small variations in inputs can be large variations of outputs, we can put some constraints on the outputs. And actually, that may not cost us much in terms of not being able to calibrate exactly, but produce much more meaningful result.

And just to be absolutely sure that our output result, our surface is smooth, we can use basis functions that smooth to begin with. So we'll use B-splines, but those will be two-dimensionals. And we'll talk a little bit more about this.

And it's not a requirement for us to have as many basis functions as we have instruments, because we can put some other constraints. Like, for example, we can put smoothness or gradient smoothness to the surface. So let's pick some relatively high but reasonable number of functions-- could be more than the number of input instruments-- and see what we can do.

So first of all, let's build our basis functions for the surface. So this slide we already just saw, that we selected to use B-splines, which is very convenient to work with. This is a one-dimensional. This is the way we build them, typically.

So we use the Cox-de Boor recursion formula. You start from linear, right? Then you apply that formula. You transition to the basis set of the second order. And then the next iteration, you have the third order. And those ones will be built.

So now, if you take those basis functions in one dimension, and same basis functions in the other dimension, and then you compute the kind of 10s of products, like you multiply them one by one, then you get basis functions with shapes like. Which means that no matter what we do, right? Like because every basis function makes sense. then any [INAUDIBLE] combinations will also be good enough.

So to formulate the problem is very simple. So we're saying, OK, the quotes produced by our model should be close enough to what's observed in the market, with some weights again. But we don't require any more that those are calibrated exactly.

We are going to put some penalty function to the change of the volatility surface. And we're going to put some penalty function to the volatility surface itself. So those are vectors, right? And L1 and L2 are matrices.

So just to give you an example what those matrices should be, like if you are talking about smoothness, if you want to penalize the gradient of the vector, right, then the matrix will consist of rows of one and following minus 1. So what you're saying, OK, if I want to penalize the difference between this and the next. And you do that for every element, OK? And the penalty kind of consists of all penalties that you have.

So here we just formulate our problem right? So we want once-- because we've had the Jacobian-- we want to price things close enough. And there are two penalty terms here with the different regularization parameters.

So once we have this, we can just, using linear algebra, the solution is defined here. And this is resulting calibration, which we see is nice and smooth.

So if we take the analysis of the calibration inverse problem, let's do that using our linear algebra tools to understand where the problem is coming from. OK, so  $A$  is a matrix, translates our model parameters to market observables. And there is some error there--  $\epsilon$ , OK? So you can see that your solution is a linear combination of singular values divided by the singular values, OK?

So if your values are high, OK, then that's not a problem. The problem is that once you get very small, singular values, OK, the derivation of VIs can result in the large deviation of your reconstructed result. And that's when you have the problem. So this is described on this slide. So "ill-posed" is that small noise may be significantly amplified by small, singular values.

And if you have a problem when you don't know how good it is, and whether you can trust it or not, so it's a very standard approach, you compute the condition number, which is the ratio of the maximum to minimum singular values. And if the number is high, which means that there are some very insignificant modes in your input data that can cause substantial changes in your output. And if you know that, it's not comforting right? So if that mode actually doesn't present in reality, then that's fine. But there is no guarantee, right? See, if that happens, then your model basically blows up.

And that slide displays exactly that noiseless situation, where it looks like if you don't have any noise and if your model is perfect, then you're always able to calibrate exactly to the market observables. But it's never the case right? So there's always uncertainty to the numbers that you're calibrating to. And your model is not always

perfect.

So very standard technique for that particular problem is the taking of regularization, which, when you solve your ill-posed problem, as trying to minimize  $x$  minus  $y$ , you add some penalty to the amplitude to your solution. Which essentially saying, OK, give me something reasonable, but something that's not blown-up.

If you go through this linear algebra, to see how that lambda parameter in the Tikhonov regularization affects the weights of the SVG kind of representation of your solution, we now see that small, single values is no longer a problem, just because we're not dividing by the small number, but actually we are kind of limited by some regularization parameter.

And typically, when you apply that regularization, your model no longer gives you a perfect match, right? But the result is much more meaningful, and more stable.

Another approach to the problem is-- and before we go to that slide. So kind of taking off regularization we used for surface calibration. Here, a standard Tikhonov regularization is something that you just penalize the amplitude of the solution itself. But it doesn't have to be the amplitude of the solution. It can be some linear combination of your solution. And in terms of calibrating of volatility surface, we didn't apply penalty to the reconstructed volatility, but we say it's not that the amplitude of the solution that we don't like. We don't like non-smoothness. So let's penalize the derivatives of the surface into different angles.

Another approach would be to use a truncated SVD, where we say, OK, so we did our single-value decomposition. We're looking at the spectrum of single values, and we find that some numbers look nice and large, some very small. And then we just skip the small ones. It's very similar to the PCA approach for the risk management that we saw before, where we just selected five principal components and we ignored the rest.

As a result, the model is much more robust. And by doing this, we essentially

truncate the null space of the model. If you're familiar with this, it's basically the space that has very small single values.

So what regularized models gives you is that improved stability. It's absolutely essential for ill-conditioned problems. And it's a more realistic and meaningful result at the expense of some beauty to fit exactly the data, but that's something that is quite often acceptable.

It might cause a biased solution, meaning that your solution again may not be exact. It might be biased towards some better result. For example, if you apply smoothness constraint, the solution would kind of assume a little bit more smoother result than it actually is. But that's acceptable.

And the bias, again, can be minimized by reasonable selection of what quantity you actually don't like. Again, you can say, oh, during calibration of our volatilities of our surface we could have said, you know what? Let's just open the textbook and see what's the regularization. OK, we find Tikhonov. We start to penalize the amplitude. Then the result won't be good.

So we need to think about and say, OK, what exactly we don't like. Like, for example, is like absolutely flat-volatility surface good for us? And we'll find, yeah, that that's actually fine. Then, if we said that then penalizing the amplitude doesn't make sense, so we need to penalize something that is a deviation from that perfect flat solution. Or to be more kind of precise, we penalize like the derivative in different directions.

So this kind of concludes my presentation today. And there are some useful links if you want to get more information.

Thank you. Any question?

**AUDIENCE:**

Yes. So regarding the techniques that you use for fitting function that you are using spline techniques. What other techniques-- is the spline the best technique you use?



**IVAN MASYUKOV:** Well, spline is, yes. So a spline or interpolation is the same. So we're always talking about interpolation. So you have some limited number of inputs, and you want to draw in between. So there are just two words for this, which is a interpolation-- or spline-- which I consider to be the same thing in general.

**AUDIENCE:** I have a question about the interpolation graph that you had where the following was very smooth. When you, as an expert in this, look at that graph and see, I guess, just some odd shapes at certain parts of the curve, how do you interpret that? And do you assess that that's a feature of the current term market liquidity conditions, or possibly just a mathematical--

**IVAN MASYUKOV:** Well, first of all, I mean, that grid is done like for-- every element is a three-month. But what's traded on the market? Like, typical maturities are three-month, maybe half-year, one year, two year, five year, 10 year. So we should kind of rescale it in a logarithmic scale, or-- you know what I'm talking about? And then, if you do that, then this peak doesn't look a peak anymore.

So the reason why it looks like a feature to you is because it's quite sharper than this guy, right? But that's because you have many more detailed instruments there compared to here. And that's the reason why we selected our basis functions like this. So we selected more node density in the front, just because there are more instruments in the front than at the end.

So we want our spline to be more detailed at the beginning, and kind of just nice and smooth at the end. So that's why those basis functions-- which correspond very well to actual instruments that we have in the portfolio-- can produce, first of all, you see this spike here too, right? So is it like you can compare this guy to this guy. But essentially, it's just they have similar magnitude, but we don't have enough instruments in this area to support any sharper features. So I don't see any problems with this graph.

But traders, because they look at this every day, right? And then they calibrate, and

they see a feature, and then they immediately kind of trying to think. OK, if you see something that you know is not typical-- and that sense of typical/not typical comes with years of experience-- then they try to arbitrage this. Because if there is a spike in the surface, it's very likely that it will disappear soon.

**AUDIENCE:** So if this is the model for volatility, that's used in modeling for swaptions, right?

**IVAN MASYUKOV:** Yeah.

**AUDIENCE:** So if you were to actually try to go about making a trade out of some discrepancy that you see, can you kind of describe how you'd do that with swaption? Do you use basically like regular options?

**IVAN MASYUKOV:** I don't have the screen here. But essentially what traders do, right? So this is, I mean that calibrator we actually use. And it's a real time calibrator. The reason why it can be real time is because there is just simple linear algebra there. Most of stuff, like a transpose a can be precalculated. So it can be real time. So they can see this volatility surface moving while we connect to actual market data.

And once they see there is anomaly on the market, there is something traded which they believe is wrong, OK, they're just make an advantage of that. They just make a trade, that would be swaption, for example, or maybe some other trade more exotic than swaptions. But again, having like the dependence on the particular instruments, which kind of would be express your position that this will change like in a day or so. So that's exactly how the desk makes money.

So we call it relative value analysis. So if we have a tool like that, you have a model. You have your input instruments, and you have some regular, as in terms of it could be smoothness, it could be a PCA, it could be combination of PCA, right? But then that additional information allows you to find anomalies in the market. Once you find those anomalies, you can take advantage of them-- provided that your model is robust enough.

And if you are saying, well, I am kind of doing well, and I'm calibrate well with just some smoothness assumption about smoothness of the forward rates, there is

nothing more fundamental than this. So if your model is based on fundamental principle, you can expect that it will be more stable in the future, rather than PCA. Because for PCA, you just kind of say, OK, I took the time interval. I kind of did my regression analysis, whatever. But that doesn't mean that the market will continue to do the same in the future.

**AUDIENCE:** I have a question. [INAUDIBLE] marketer. And then we would try to price the bond at that premium. You mentioned that actually bond is the most liquid instrument in the market. So why not you do the area around by inverse bond to derive the [INAUDIBLE] vector from a bond [INAUDIBLE].

**IVAN MASYUKOV:** Well, that's a very good question. So basically we could have done that, OK. And some firms do that.

The problem is that with swaps, we kind of have those swaps today, they kind of roll every day, OK? They like, the swap today starts today, swap tomorrow starts tomorrow, things like that. But bonds do not. OK so they basically, there is like on-run bond, which is the most liquid one, which once the new ten-year bond is issued, then that bond becomes off-the-run. It's still traded, but then everyone switches to on-the-run.

So you don't have a nice continuous spectrum of bonds. You have kind of concentrations between on-the-runs and off-the-runs. And if you want to draw the curve for all of them, you typically cannot do the perfect fit. You kind of need to do the least squares. So it's just more convenient to do it in the swap.

But once we build the swap curve, swap trader typically-- I should say always use bonds for hedging, just because bonds are much more liquid. Then we project bonds to the swap curve rather than swaps to the bond curve, which is hard to build.

**AUDIENCE:** So in this case, when they switch from on-the-run, off-the-run, [INAUDIBLE]?

**IVAN MASYUKOV:** Yes. So your curve won't be stable, just because those roll effects-- we call those "roll effects"-- which means that there is something substantial changes on the market. And there may be such a big demand for this new bond on the market, that will make your curve, that won't look nice.

So there are also traders that just trade bonds. And those typically don't have curves. They rely on some PCA models, or some other things.

**PROFESSOR:** Thanks again.

**IVAN MASYUKOV:** Thank you.