[SQUEAKING]

[RUSTLING]

[CLICKING]

TOBIAS SALZ: OK, let me start here. So thanks for allowing me to be back one more time. Today I'm going to talk about empirical models of auctions and hopefully fill in a little bit what you have already learned from Glenn.

So this is the roadmap. The goal is to introduce you a little bit to how the thinking has evolved on the empirics of auctions and then bring you to what is the state of the art still, which is Guerre, Perrigne, and Vuong, that has really had a huge influence in the way that people analyze auctions empirically.

The gist of all of this is that sometimes it's hard to compute equilibrium for auctions, but it turns out that under the right conditions, it's actually really easy to estimate models for auctions. And I'll show you how to do that. And then we're going to talk about two different applications that are quite interesting. So what is the relevance? Why do we want to look at auctions?

So for one, we are looking at, as IO commonly says, markets where firms or players have market power. And a seller, let's say, in an auction setting has, of course, market power. When you think about the problem of setting the reserve price, it looks almost like a monopoly pricing problem. So it's kind of squarely within the topics that we're interested in. There are estimates that about 10% of GDP is contracted through auctions and many different designs. So it's important to understand how this is being done.

Very often, the government is involved as either a buyer or a seller. If you think, for instance, about spectrum auctions, if you think about highway procurement projects, the government gives projects to construction firms. Other settings, electricity auctions, treasury auctions, and so on, all of these have, of course, direct policy implications because let's say you want to procure a construction, road construction at the minimal cost to taxpayers, you want to figure out what is the right mechanism to use.

So there is this tight link between the way we analyze these kinds of settings and what we can actually do as policy makers, such as setting the reserve price, thinking about which mechanism to use.

More from a research point of view, auctions have been quite fruitful in analyzing collusion. So oftentimes, it's a bit easier to get traction on collusion than, let's say, in repeated games. And there are many nice studies, some of which we will discuss in [14.]273.

What is also really nice about auctions is that the rules of the game are very defined. We know what the strategies are. We know who the players are. And we actually see a lot of things. So if you think, for instance, about something like multi-unit auctions, like let's say in electricity, we actually see the whole supply schedule that a firm bids. That's something that we spend a lot of time-- supply and demand curves usually spend a lot of time backing out, but in some settings, we actually directly observe at what prices a firm wants to supply.

And so that gives us a lot of empirical traction. It also means that whatever model we write down is oftentimes very close to the model that firms are actually playing. So the link here is very tight. If you think, let's say, about entry games or other settings-- if you think about what Amazon can do, Amazon can do a million different things. But once we narrow it down to a specific auction setting, we kind of know what the rules of this auction are, how people are bidding.

And then the participants are oftentimes quite sophisticated. So take, for instance, the example of spectrum auctions. So I mean, these are very, very expensive bits that, let's say, firms like AT&T or T-Mobile put in. And what they're typically doing is they hire a firm, like Paul Milgrom's firm. And then Paul Milgrom devises their bidding strategy. And so suffice it to say that they're probably bidding pretty sophisticated in that case.

So that means that when we write down a model where firms are optimizing, maybe actually getting kind of close to what's actually happening. Always the goal is to recover a distribution of bidder valuations. So really, to the extent that we think of valuations, unobserved valuations or unobserved, let's say, costs in the procurement setting as the structural error term, we're really interested in backing out this one distribution, the structural error term.

So let me just very briefly recap single-unit auctions. I know you have seen this probably multiple times at this stage, but I will refer back to it a few times. So we have n bidders, indexed by i. The evaluations are denoted by Vi. They privately observe a signal that summarizes the information that they hold. And they're bidding a bid Bi.

And typically the convention is that we refer to random variables as upper case letters and realizations as lower case letters. We are staying today in the IPV paradigm. That means that signals are distributed i.i.d. And we also stick to the private value paradigm. That means that I can learn from my signal about what signals other players have drawn. So this is a specific case where I only need to look at my signal, and that's all I need to know.

OK, again, this is recap, I notice. But because I come back to this, I want to walk you through this once more. So we can solve in the first price auction case for the unique symmetric Bayesian Nash equilibrium. Bidders are maximizing their valuation minus their bid times the probability of winning. And this is the distribution of valuation. So we want to map back to this. And so we take the inverse of the symmetric bidding function, mapping back from the bit to the valuation.

You can take the first order condition. We get a differential equation that you have seen before. Later on, when I discuss GPV, I will come back to this. And here, you get with a boundary condition the closed-form solution for the bidding function.

So let me just comment, first of all, that this doesn't have a great economic interpretation, but it's equal to the expected value of the valuation of the other remaining bidders, given that my valuation is higher than theirs. So that's what this is equal to. So that's a bit more interpretable.

The other thing from an empirical point of view, if you stare at this expression here, this is an integral over a CDF. And for most common distribution functions, and especially if we do not impose any parametric restrictions, that does not have a closed form solution. That means that from an empirical point of view, this is kind of a messy thing here because we have to numerically integrate this. And so I come back why that's not a great thing and how to possibly sidestep this.

So here is the solution with a reserve price. OK, so let me come to the first paper that I wanted to discuss today, Laffont, Ossard, and Vuong. This is one of the very early empirical auction papers. It has some really nice ideas. It's not really the way the literature has gone, but I think it helps the way to understand how the thinking has evolved.

So again, we're in the symmetric IPV setting. These are descending auctions with a reserve price. These kinds of auctions are oftentimes used for perishables, things that have to be sold quickly. Because as soon as somebody stops-- you only need one bit to stop the auction. The setting here is a little town in the south of France where they're selling baskets of eggplants in a wholesale market. So there are sellers that buy these baskets and then go to a retail market and sell those.

The imposing log normal distribution on the valuations. And the goal is to estimate the parameters of this log normal distribution. And they have data from 81 auctions with the same set of bidders in each of the auctions. OK, so how can we make progress on estimating this distribution of valuations?

So one idea is to use the-- we have just seen a solution for first-price, sealed bid auction. One idea is to use the strategic equivalence between sealed bid, first-price auctions-- with first-price, sealed bid auction and use maximum likelihood. So let's see how far this gets us.

So I want to maximize some likelihood function, where f is given by this log normal as the density of the log normal distribution and the vector of bids that are observed. And I know here, with a reserve price built in, I know that this is my closed-form solution to the bidding function.

So now, if you think about how to actually do this, how would you do this? For a given parameter guess, I know what these objects are. What I observe are the bids. But I want to evaluate the likelihood, of course, at the valuations, which I do not observe.

So what I need to do is for a given guess of parameters, I need to compute this thing. And I need to find the v under this expression that makes this equal to the bid that I observe. So I basically need to conduct a line search, where I repeatedly guess v's until this expression is equal to my b.

And while I'm doing this, I need to repeatedly evaluate this thing that I told you is not super straightforward to compute because it has this integral over a CDF. So you need to numerically integrate this to support that integration limits are changing. That makes all of this not so easy.

So you probably get the sense that this doesn't work great. In the early literature by papers by Harry Parr and others, they have done this. But this is no longer really the way people estimate auctions these days. It's computationally costly, as I hope I made clear here. Another issue is that we actually only observed the winning bid here.

So what we would actually have to do is we have to first get the expression for the likelihood of the winning bid, which is something we can, of course, do. But there's an extra, more fundamental problem that the support of the distribution of the winning bid depends on the parameters, and in particular on the max of the largest out of N minus 1 draws on the distribution of valuations and the reserve price.

And that violates the regularity conditions of maximum likelihood. So maximum likelihood in this case is not a consistent estimator. So there again, there's early work by Harry Paarsch that tries to find ways around this. But as you will see, there are much better ways to go about this.

OK, so here's another idea. So instead of having to map back from the expression for my bid to my valuation and then evaluate the likelihood, what I could instead do is I could simulate the object that I observe. So I see a bid, and with the closed form expression for my bidding function, I can simulate the integral and match this integral, the expected winning bid, to the observed winning bid.

OK, so here is the expression for the expected winning bid, which integrates out over this bidding function and here uses the expression for the density of the highest valuation. And so when you plug this in, you can simplify this a little bit, and you get an expression that looks like this here.

And typically, complicated integrals, one question you should always ask yourself is, instead of computing this integral exactly, do I somehow want to simulate this? Because simulation is often oftentimes easier. In BLP, that's something that we typically do for the random coefficient. We take a bunch of random draws and then integrate out. So that's something one could do.

It's still somewhat cumbersome. I mean, you have to take random draws from the log normal. That's not easy with-- that's actually very easy with modern programming languages. But then you still have to-- for every draw of the evaluations, I have to compute this expression here over which I take an average. And here still I have to compute this integral over a CDF. So it's already much better than what we had before, but it's still somewhat cumbersome.

Any questions so far?

So here's a clever insight that Laffont, Ossard, and Vuong had. We stick with simulation, but we can use the revenue equivalence theorem. So I've seen that Glenn has talked about this. So I'm just repeating this here. So we're assuming a risk-neutral IPV setting with atomless signal support.

We know that in this case, any mechanism which is efficient in that it awards the object to the bidder with the highest signal or with the lowest cost and leaves the weakest bidder with zero surplus, it's the same expected revenue. And so how can we use this?

We know that using this result, for this specific sets of assumptions, the expected bid that we get from this closed-form solution for the first-price sealed bid auction, here using again the expectation over the valuations of the strongest bidder, so the highest order statistic and taking into account the reserve price, that the expected revenue from those bids has to be equal to the expected revenue from a second-price auction within the same setting.

And remember, in a second-price auction, bidders' weakly dominant strategy is to bid their valuations. And so here we can basically make use of the fact that I no longer have to compute any complicated bidding function. I can directly take an expectation over valuations.

So how would one then do this? This is actually really simple. So you generate-- you have your log normal. You generate some-- for a given simulation draw, you generate a vector of evaluations. You take the second highest one and set the winning bid as the maximum of that realization and the reserve price. You do this a bunch of times, and you get an expectation for the winning bid.

You do this for every parameter guess and you minimize the squared deviation. So this is nonlinear least squares from the observed winning bids. So this is to show that sometimes you can use the theory to make your life really easy. In this case, this clever insight that instead of estimating this with a first-price auction, we just use a second-price auction instead. And this makes our life much, much easier.

So let me comment on this a little bit. So this is, as neat as this, actually not an approach that's widely used anymore in empirical work. So one reason is certainly that there are many different ways that revenue equivalence breaks-- asymmetric bidders, risk aversion, endogenous-- I mean, cost of entering an auction, you know, collusion--

So many things that we believe are important in the real world lead this result to break. And so it might be maybe assuming a little too much to rely on this result. The other reason is that people, at least in this literature, in the auction literature, have moved away from strong reliance on functional forms.

So in this paper, we just assumed a log normal distribution, and people are typically not comfortable to say from the get go, oh, this is what the distribution of valuations looks like, or this is the parametric family which in I believe it lies.

And so instead, what people are doing is they typically try to identify models, what is called nonparametrically, and place strong emphasis on formal identification results. So I want to now just have a digression of one slide to talk you through a little bit what that actually means. Yes.

AUDIENCE: Yeah, so was it the case that for each option, we're assuming that the bidders have a new valuation drawn?

TOBIAS SALZ: Yes, in this setting.

AUDIENCE: OK. And so I think at the very beginning, you said that there's the same 11 bidders throughout. And so for a given bidder, the valuation draws are correlated at all? Or are we also assuming that those are independent?

TOBIAS SALZ: In this particular application, they're assuming that those are independent. But that's a good question, that you may think that some bidders maybe have better resale opportunities, have a better retail location. And for them, their valuations are kind of systematically different over a longer time horizon.

And these kinds of asymmetries are oftentimes, in many modern applications of auction models, people try to take them into account. They make life very complicated because for asymmetric auctions, there's typically no closed-form solution for the bidding function in the first-price auction case. So it's estimation, as I will show you in two slides, is easy. But equilibrium computation is hard with asymmetries.

So just very briefly, what does it mean for a model to be identified? People use the term identification in many different ways. I'm telling you how this term is used in the auction literature. So what people have in mind here is that-- so first, we are settling on the game, and we're settling typically on a solution concept for this game, let's say, a Bayesian Nash equilibrium.

And given that, a model is identified if the joint distribution of bidder utilities and signals is uniquely determined by the joint distribution of things that I observe. Things that I observe are typically bids or bids and entry behavior, or in some settings, bids and, let's say, whether I'm drilling this oil field after the auction or not, so things that are observable to the researcher.

OK, so here's a slightly more formal way of saying this. Let m star be the true vector of functions and distributions, and let P of m denote the joint distribution of observable variables under the assumption that the data is generated under m. So think of P as basically a mapping from my model. And just to make things simple for now, think about your model is the distribution that you assume for bid as valuations. And P basically maps that with some assumption on which game is being played and what equilibrium to the data.

And we say that m star is identified in a class of models if and only if for all models in this set, whenever my model is different from m star, I would also observe something different. And so what is important here is that this-- we typically do this-- I mean, we do this under the assumption that we have infinite amount of data.

So this is not something where we're worried about small sample properties. This is really in the ideal case. Sometimes it might be rare events that tell two models apart. Here we're assuming we have these rare events that allow us to tell the two models apart.

So now with this, I can tell you what it means for something to be nonparametric. So first, a parametric model is if I restrict m to be a subset of a finite dimensional space. Let's say I'm searching over the set of normal distributions or exponential distributions. Sometimes one parameter may be enough. Or I could even allow for, let's say, a search over different kinds of parametric distributions that would still be parametric or semiparametric.

If m, however, is not a subset of a finite dimensional space, we say that it's semiparametric if a subset of functions lies in the finite dimensional space and nonparametric if none of them lies in a finite dimensional space. So thing about nonparametric basically means that you can have any sort of weird density of valuations. It could have different modes. You're not placing any restrictions on it.

So one more word on this. So how do we actually go about thinking about identification? So the typical approach is what is called the constructive approach, is to try and express the things that I do not know in terms of things that I observe. Because then it's crystal clear that what I'm looking for-- the primitive that I'm looking for can be expressed in terms of data directly.

So just very, very simple example for this. Think about a second-price auction. So again, for us, we're typically interested in learning the distribution of valuations. And let's say we assume that the bidders here are playing the weakly dominant strategy of bidding their own valuation.

I hope that-- you see this, right? This is basically almost immediate. I know that, in this case, my bids are equal to the valuations. This means I directly observe the valuations. So it's trivially identified because I can express-- every bid as valuation, I have a direct correspondence to that bidder's bid.

So what, however, if only transaction prices are observed? OK, so now I can do this. I can't directly express the unknown thing, my v in terms of the known thing b. So here the following result comes in handy. So first, as you know, the winning bid-- sorry, the transaction price is given by the second highest valuation.

So what I do observe is the second highest order statistic. And what we typically-- not always, but what we typically assume in auctions is that what we also know is the number of bidders. This is empirically not always crystal clear that that's really something that's observed without measurement error or precisely observed. But that's typically an assumption.

So I know the number of bidders. I also know the second highest order statistic. I do not know f of v directly. I only know the second highest order statistic. So can I make progress here? The answer is yes because of this result. I think Glenn in his lecture had also an expression for order statistics. This is yet a different expression for order statistics.

It makes it pretty clear why, from any order statistic, we can identify the underlying distribution. And that's because if you stare at this expression here, everything on the right-hand side here is known except for f of v. I know N. I know which order statistic I observe. And this is an integral that's strictly increasing in f of v. So that means I can pointwise invert this thing.

I know this here pointwise. And so for every point in the support, I can basically map this to an f of an underlying f of v. So again, what this shows is that even though we do not observe all the bids, we can still make progress.

And so what this, of course, also means is that any order statistic identifies any other order statistic. So I hope this gives you some sense of how people normally approach this. So again, here the idea was to-- just coming back to this once more-- this is theta, this is the thing that I'm looking for, this is a known function that I can invert. So I can express f of v in terms of theta or known functions of theta.

OK, so this is really the landmark paper in this literature, Guerre, Perrigne, and Vuong. Think of this as a paper for auctions as BLP is for demand. This has really created a big literature. The idea of the paper is, again, similar to these easy steps that I've just walked you through, is that we can directly express the unobserved valuations in terms of things that we observe.

And so I'm not going through all the details here, but this also works if we do only observe the winning bid in most cases. So I promised you I'd come back once more to this first order condition here from our first-price auction sealed bidding.

So if you think about this, it looks pretty daunting, What we observe are the bids. And we know the number of bidders, but we do not know v, we do not know the distribution of valuations, and we do not know the derivative of the bidding function.

You still have the insight that GPV had, which is that because of the monotonicity of the bidding function, a percentile of the bid distribution has to correspond to a percentile of the distribution of valuations. OK, so the 20th percentile of bids that are observed in my data must-- that percentile must have been submitted by somebody who is in the 20th percentile of the distribution of valuations.

So that means that we know that G of b is equal to f of v. And we also know that G of b is actually given by this object here, G of the symmetric bidding function of v. And then you can take the derivative on both sides, and you get that G of b is equal to f of v times 1 over the derivative of the bidding function.

So now you probably see that we can make some progress here. I can basically replace f of v here by G of b times beta prime. Beta prime cancels out, and I can replace f of v by G of b. And that means that I can-- again, as I said, this is what is called a constructive approach. We can express valuations in terms of bids and distributions, CDFs and PDFs of bids that we observe.

So all of this here on the right-hand side is theta. This is the thing that we do not know. So for each bidder separately, we can figure out what is that bidder's valuation. And so this should also make it clear why this is-- automatically, it implies that it's nonparametrically identified because we don't place any restrictions. For every separate bidder, the bid is exactly rationalized by a specific valuation that I can back out point by point.

So this is how this is derived in the paper. There's an even easier way. I want to show you this as well. And once you think about it this way, it's sort of ex-post obvious that you can do this. But oftentimes with good papers, they're ex-post obvious at least. So there's another way of deriving this. This, by the way, a trick that's used a lot in IO. So that's something that generally you should remember.

So we observed the bids in the data, right? And we are already committed to our assumption that this is coming from some equilibrium, So that means that we observed the equilibrium distribution of bids in the data. And at the point where we're willing to stick to this assumption, we can write down a bid as a problem instead as what essentially looks like a single-agent problem, somebody who's maximizing against this distribution of bids that's just given by data.

Because, I mean, that is the assumption that we're making-- the best responding to some equilibrium distribution of bids, and we observe what that distribution is in the data. And so this allows us to write this down in this way. And we get-- this is now really simple because we just take first order conditions and we directly arrive at this expression here after cancelling out and rearranging.

So this is an idea that's also used in repeated games and other settings where typically equilibrium computation is hard. You say, let's not even try to first build things up from primitives and then compute from primitives what the equilibrium looks like. No, we observe the equilibrium in the data, so we don't actually need to compute it, because we already have it.

So just a brief comment on how you actually do this. So people don't always follow these steps. But as you will see later on, in terms of-- just to remind you what we need to do to compute these v's, bid by bid-- we have the bids, but we still need to compute the CDF and the PDF.

What people typically do is-- of course, the CDF is really easy. They use kernels for the densities. Then there are some practical questions like how to choose a bandwidth, and sometimes you want to trim in the tails because observations, they are noisy. And this term, as you see, because we're dividing by the densities, can easily explode if the density becomes very, very small. And so people sometimes throw away some observations in the tail.

Standard errors are typically in practice computed via bootstrap or functions of interest such as the reserve price counterfactual objects that you get out of these estimates. Any questions? Yes.

AUDIENCE: Yeah, so for this one, this specific example, we still have to go back from observing the highest bid to the density of the whole bid distribution, right?

TOBIAS SALZ: You may have to. I mean, it depends on your setting.

AUDIENCE: We're observing only the winning bids, right?

TOBIAS SALZ: Right. And then you can, of course, map from the winning bid distribution to the bid distribution. And you can simulate from that bid distribution. And with that, you can then-- yeah.

OK, so let me comment on this a little bit. This is really, really simple. One of my advisor, John Asker, he once joked that his most famous paper is probably the one on bidding rings that uses GPV. And he said he could have done that paper in Excel. And I think it's a bit of an exaggeration, but this is something that is so easy. I think Excel has kernels. You can probably compute valuations in Excel if you wanted to.

And if you compare this to this process that I walked you through earlier, that's, of course, a great improvement. It has this advantage that it does not rely on functional form restrictions. And as I said, this has been a very influential paper, many, many different extensions to multi-unit auctions-- very similar ideas apply. If you have dynamics, you can still typically make a lot of progress.

I'm going to talk about Athey, Levin, and Seira and some really interesting applications of using it to quantify the damages of bidding rings. The maybe criticism not of this paper but of the literature is that, well, the IPV assumptions are really, really convenient. And that leads people oftentimes to use the IPV assumption because they know how to make progress.

And in many settings, we believe that maybe there is a common value component. Bidders do not know what they're really bidding for. And oftentimes, there are resale markets or other reasons why we believe that a common value problem may arise. So that's just something to keep in mind. Maybe one more comment on this, actually.

So you may now wonder, well, if we believe that common values are an element of common values in many settings, why hasn't the literature done more in this direction? So it turns out that if you follow the same steps that I just walked you through, we basically have mapped one distribution-- we have identified one distribution of valuations from one observed distribution of bids. That means, loosely speaking, we're just identified in terms of backing out this unobserved distribution of valuations.

In a common value setting, there are two things that you don't know. There is the distribution of underlying true valuations for the object. And then there's the distribution of signals that captures the uncertainty that-- with which bidders are forecasting what the valuation is going to be. And these are two unobserved objects, but you still only observe one distribution of bids. So you're fundamentally underidentified in these settings.

And so that's why you see many often-- even when it's kind of a-- the people sticking to the IPV assumption.

So let me now talk about a really interesting application by Athey, Levin, and Seira. The question that the paper asks is, should we use sealed bid or open auction format?

So just to remind you, in some instances, the theory tells us it shouldn't matter. Because of revenue equivalence, the seller should be equally well off. But, again, there are many things that we believe are very plausible to occur in the real world, such as bidder asymmetries, collusion, risk aversion, and so on, that may break this result.

So this is what they're basically asking. Well, if we actually test this in the real world, what do we find? And they look at this in a setting that-- you see a lot of timber applications in the early auction literature that there were some great data available from the government. This turned out to be a fairly large industry. If you're accounting for both the logging and the milling, it's about $100 billion industry-- $100 billion a year industry.

And again, there's this tight link between policy and the setting and the game that we are estimating because the government typically puts up tracts of land up for auction. And then loggers or mills are bidding on these tracts. So the way this works is that they inspect or cruise this tract and look at what is there. And then tracts of lands are auctioned off.

And they have data from what they call the northern regions, the Idaho-Montana border and California. What is really, really nice about this setting is that the US Forest Service uses both open and sealed bidding. And in some instances, even randomly so, it appears.

The other thing that makes this interesting but also complicates things is that there has been a widespread suspicion, believed by not just theorists but by practitioners in this industry, that open auctions foster collusion. Actually, Congress at some point passed a law that says that all these tracts should be should be auctioned off by sealed bidding. And then, later on, they started making exceptions.

And it turns out that the large bidders in this industry really favored the open auction format. And you can guess why. Maybe it's a bit easier to monitor who is doing what. And that makes it easier for bidders to collude. So collusion is something that has to be kind of accounted for here.

So what they do is they write down a model that captures some salient features of this industry. One salient feature that I told you about is that the government uses both auction formats. So they actually have to think about equilibrium behavior, both in open and sealed bid auctions.

The other really salient features is that they're loggers and mills. Mills are basically bidders that have some production capacity of their own, so they can refine and work with the timber. Whereas loggers, they actually first have to find a buyer in the form of a mill. So they have to pay-- they have to take an extra cut because they have to sell it to mills first.

And so what they're assuming is that loggers are weaker bidders in this setting. And precisely, they assume that the hazard rate of the value of a mill's valuations dominates the ones of loggers.

This is a setting where there is variation how many people are bidding. So they're accounting for endogenous entry. And they're searching for an equilibrium in type symmetric entry and bidding strategies. So loggers and mills-- so conditioning on loggers, they all use the same entry probability and bid once they're entering according to the same bidding function, and same for Mills.

Two more assumptions that they're making is, first of all, mills are not just stronger. They're basically sufficiently strong so that your mills are always entering first. So basically, the profits of mills with an extra mill entering are still larger than the profits of loggers if there is no such additional entrant.

And they make this assumption to reduce the number of equilibrium. And the second assumption is that only mills are colluding. So to the extent that there may be collusive equilibria, they're only coming from mills colluding with each other, maybe because these are players that have been in the market for longer. And for them, it's easier to meet repeatedly and organize this collusion.

So they get a number of predictions out of the model. So the first one is that they get a unique type symmetric entry equilibrium, where either no loggers enter or all mills enter. So in other words, what we cannot have is that we see a logger entering, but not all of the mills enter.

And that's because of this assumption that profits are sufficiently higher to overcome an additional bidder in the auction. So then the key question of interest, to remind you, is, should the government be using opened or sealed bidding in these auctions? And here are some comparisons of these two auction formats.

The model tells us that in open auctions relative to sealed auctions, loggers are less likely to enter, mills are more likely to enter, and it is less likely that a logger will win.

So just to explain this a little bit, this goes back to result in the auction literature that's well known by Maskin and Riley, which is that if you have asymmetric bidders, the weaker bidders tend to bid more aggressively. So in this case, they would be bidding higher markups over their valuations than stronger bidders.

And so this means that sometimes-- first of all, it means that sometimes the auction can be inefficient. It doesn't actually award the object to the bidder with the highest valuation because the weaker bidders are bidding a higher markup. And it also means that the sealed auction format is more attractive to weaker bidders because in the open auction format, everybody just sticks in up to their valuation. And it's going to be efficient, and, more likely than not, because of the assumption that mills are stronger, a mill is going to win.

So that's basically-- the explanation for these comparative static results is this result by Maskin and Riley.

Lastly, they get some other competitive static by comparing in the open auction format. So they only consider collusion in the open auction. In the open auction format, comparing it to-- comparing collusive and noncollusive equilibria, and so here it's not too surprising that if only mills are colluding-- and remember, they're all antisymmetric-- so with some probability, each mill is going to benefit from collusion and being designated--

Sorry, one thing I forgot to mention, I should say, is what they assume is perfect collusion. So the mill with the highest valuation is going to be designated the winner. And all the other mills are not bidding to keep the price for the winning mill low. So because mills are antisymmetric, there's always some chance benefit from collusion. That means that collusive equilibrium make this even more attractive for mills. So mills are going to be more likely enter and crowding out loggers.

So now we have some predictions from the model. So again, to remind you what's really nice in this setting is that the government, sometimes seemingly random, designates tracts either for sealed or for open auction bidding. And they run this regression here where they're regressing outcomes of interest on a sealed indicator plus a bunch of controls, including in some regressions the number of bidders that are in the auction.

And they do this separately by the northern regions and in California. The reason that this matters is that the northern regions are-- there's more suspicion that there's collusive behavior here. And the other reason is that those are actually-- they claim that those are more likely to be randomly assigned. But that's just an aside.

So what they find is consistent with what the model predicts is that in the sealed bidding format, more loggers are entering. There is no significant effect on mill entry. And so the fraction of loggers in the auction goes up in the sealed bidding format. Loggers are more likely to win, but this is not-- these effects are not significant.

And the prices in the northern regions are significantly higher, both when accounting for the number of bidders and even-- sorry, without accounting for the number of bidders, you get more entrants. We know that more entrants are good for sellers. But even conditioning on the number of bidders, they find that the sealed bidding format leads to higher prices.

Largely similar predictions for California, with the exception that there is no effect on prices here from the sealed bidding format. OK, so in other words, they find support for the predictions of the model, that the sealed bidding favors weaker bidders, leads to more entry of those bidders, and at least, in some settings, the sealed bidding format seems to benefit the seller.

The second takeaway is that maybe there's something strange going on in the northern sales because the sealed bidding format leads to higher prices, perhaps because it shuts down this collusion channel.

So with this in mind, what they do is they now want to quantify the model. And they follow the steps of GPV. I'll not go through the details here. So they need to do this for asymmetric bidders. And they deviate slightly from what I told you people would normally recommend that you do to estimate a kernel. They actually estimate a parametric function for bids. But these are just details.

One thing that they do that I do want to mention-- although I'm not talking about the details here; we're going to do this in [14.]273-- is you could be in many settings worried about common values. But yet another thing you could be worried about is that there's something that all the bidders are actually observing. So they do not have any uncertainty about this but the econometrician does not observe.

So maybe there's some tracts here in the setting where, I don't know, there's a steep incline or something that makes it much more difficult to log the trees. The bidders all know this, but this might not necessarily be recorded in our data. So that means that you find that the bids from one auction to another are correlated.

And if you do not account for the fact that the bidders all observed this and take this into account, you might be overstating the allocative efficiency of auctions because you load all this heterogeneity as cross-bidder heterogeneity. You're not accounting for the fact that some auctions look really good, some auctions are really bad, and the average prices in each auction are accounting for that. Yes.

AUDIENCE: What are the timing of the auctions for the tracts? Are they occurring simultaneously? Is there kind of like a continuous time aspect that the tracts to open up in the auction? So I guess the question is, how much does it become a multi-unit auction versus not?

TOBIAS SALZ: I mean, so I don't know exactly how the timing works, to be honest. But there are many auctions. And this goes-- there are potential dynamics that result from that. So I think Susan and others they have or maybe Pat Bajari, they have a paper on capacity constraints of bidders or other dynamic things that may matter. So they occur frequently enough that these kinds of things might potentially be relevant. Yeah. I hope that answers your question.

Where was I? OK, I was talking about this unobserved auction-specific heterogeneity. Again, I might see that there's a tract with low-quality wood and every bidder appropriately prices this in. And so there's basically a way to account for this that comes out of the nonlinear measurement error literature. The key paper here is the one by Li and Vuong. And this has been extended to asymmetric auctions by Krasnokutskaya.

So this basically allows to account for the fact that there are systematic differences across auctions that bidders do not account for. And that, again, matters because otherwise you would say that this heterogeneity is across bidders. And then that would give us oftentimes very, very wrong estimates for, let's say, the efficiency of an auction.

They estimate optimal entry-- and something really neat that they do that I want to discuss a little bit is they estimate only based on sealed bidding using GPV to have the open-auction format. They have the sealed bidding format. They only estimate based on sealed bidding, but they generate predictions for both auction formats. So that is a really strong test if you think about it of the model.

You're estimating the valuations based on sealed bidding, and you generate predictions for the open auction. So let's see kind of how well they're doing. So here is-- here's a table that compares separately for the northern sales in California, the out-of-sample performance of their models. So they have some holdout auctions. And they simulate outcomes for this holdout auction from their model and compare it to what actually happened in those auctions.

And then they do this for a sealed bid. So holding the mechanism on which they estimate constant, and they also do this for the open auction format, which they actually never use for estimation.

So what they find here is you can go through this-- so the big picture is for sealed bids, the model does really, really well, as we would typically expect. If you think about-- if you think back to what we did in GPV, we rationalized each bid exactly, by finding a valuation that matches each bid exactly. So you would actually think that if you follow exactly all the steps, then your model should predict exactly what is happening.

Here, there's some things that deviate a little bit from this. They put a parametric distribution on the bid, so that might-- that could introduce some prediction error. But what they basically find is that the average bid is very close to what is predicted.

And that's true whether or not you account for entry. Same for the average revenue is very close to what is predicted. And your average logger entry is almost exactly-- I mean, is exactly spot on. And just very briefly, if you look at sealed bids for California, the same is true. So logger entry is predicted perfectly. You get very close values for average bids. And I guess average revenue in California is somewhat overpredicted.

So let me actually stick to this slide for a second. For the open auction format, and I find this amazing, they also predict, at least for the case of California, the average sales price and entry very, very closely.

So again, this is a pretty-- I haven't seen this done in many papers. This is a pretty strong test of the model, that you take it to a different setting and then see how well it performs there.

What they then do is because there was suspicion of collusion in the northern sales, they also generate for the open auction format predictions from a collusive model-- the way they generate a collusive model is to take a simulation, draw the highest valuation bidder is going to be designated the winner. The other mills are not-- are bidding zero.

And so that leads to much, much lower prices. And in fact, so much so that they, even for the northern sales, where the collusion was potentially an issue, underpredict the average sales price by about a third. But they do also reject the competitive model for the case of the northern sales.

So in this instance, they're rejecting both models. And so what they then argue is that maybe what best explains the data is a moderate form of collusion, where some of the mills are colluding or some of the mills are colluding some of the time.

So kind of my-- my view on this paper, this is a very complete and well-done paper. In fact, it does so many things that I didn't have time to talk about all of them. But it has taken a pretty big bite out of the apple. Many following papers do focus on one aspect, collusion or entry. And here, they do many of these things together.

And it's basically theory testing plus a setting where you can use some random variation to be pretty confident that the model that you're selecting is the right one, plus then the structural exercise. So this is pretty good blueprint for a job market paper because it has kind of all these different things in there.

So maybe some things that one could complain about, again, is this a private value setting? Certainly there are resale markets for the timber. And so that, in itself, introduces a common value aspect because everybody can sell at that resale market price. And maybe there's some uncertainty about the quality of the timber or how difficult it is to get it.

And then the other question is, can we attribute part of this effect that we see that the sealed bidding format generates much higher revenue to risk aversion? And so I want to use my last-- I'm going a bit fast today.

But I'm going to use my last minutes to talk about another paper by Yunmi Kong, who was my classmate in grad school, which thinks about risk aversion in open and sealed bid auctions. So the setting here is a very different setting. It's the Permian Basin in New Mexico. So the Permian Basin is a big-- one of the most productive oil fields in the world that spans New Mexico and the west of Texas.

And here we have many smaller firms, oftentimes operated by just a few people, that lease land from the government and that rent equipment-- here you see an oil rig in this setting-- that rent equipment, and basically, through these auctions from the New Mexico State and Land Lease Office, get the land, and try their luck and drill for oil.

And so what they're bidding on is a lump sum bonus. And then for the time that they're renting, they have to pay royalties and a rental rate. What's nice is that, as in the previous paper, it seems that-- So the government uses both. Or the New Mexico State Land Office uses both open and sealed bidding again. So we can get a similar comparison as before in a very different setting.

And if you stare at the map of how these formats are spatially distributed, this is open, this is sealed, it looks pretty random. So again, we can use basically with some controls the fact that this looks almost random to look at the effect of different auction formats.

So here's what she finds. As before for the northern sales, sealed bidding leads with a bunch of controls for oil prices and differences of these tracts of land leads to about 30% higher revenue than the open auction format.

The other really interesting observation that she makes is she compares the distribution of prices across open and sealed bidding. So here are the histograms put on top of each other. And you see for the open auction format this big spike here.

So the reason this happens is that oftentimes bidder enter the open auction format, they're discovering they're the only bidder, they're bidding the reserve price. Interestingly, although the distribution of the number of bidders for the sealed bid auctions is almost identical, and there are many auctions that also only have one bidder, we do not see such a mass point.

So what this suggests is that the bidders do not know how many other bidders are going to be in the auction when they engage in sealed bidding. So they have uncertainty, and they basically want to insure themselves against competitors by bidding higher than what they would if they were the only bidder, even though with some chance they are going to be the only bidder.

So this observation together with the previous observation that the revenue for sealed bid auction seems to be higher leads Yunmi to conclude that risk aversion is probably the explanation for this.

The reason is that uncertainty is revenue neutral for the seller but only if the bidders are-- sorry, if the bidders are risk neutral, then uncertainty about the number of bidders is going to be revenue neutral for the seller. And that's not what we observe. So that's a theoretical prediction from Harstad, Kagel, and another Levin. And that's not what we observe. We see higher revenue for the seller from the sealed bid auction, which is consistent with this idea that bidders insuring themselves against competition in the sealed bid auctions by bidding higher.

So I think this is a clever detective work to make the case that even when companies are competing against each other, where we typically assume risk neutrality, at least in this setting, where many of the companies are small, it seems to be that risk aversion plays a big role and leads to higher revenue for the seller and sealed bid auctions.

And she then goes on to show that the model, a model with risk aversion, where now you not just have to identify the underlying distribution of valuations, you also need to identify the level of risk aversion and privacy-- like I said we just identified in the standard case-- so here she basically uses both the open bidding format and the sealed bidding format together to show that you can identify the distribution of valuations of bidders as well as the level of risk aversion.

So it's a nice paper that's in *RAND.* OK, ending early. That's all I have for today. I hope I see many of you next semester in [14.]273, where we dive deeper into all of these topics.