[SQUEAKING]

[RUSTLING]

[CLICKING]

TOBIAS SALZ: So let's get started. So welcome, everyone. I'm very excited to be here for a few lectures this semester. I thought I start by introducing myself. I haven't met many of you, in fact. So I was on sabbatical last year. Many of you are first year. So glad to meet all of you.

So my name is Tobias Salz. I sit in E-52 on the fourth floor. As you have seen on the syllabus, my office hours are on Thursdays from 1:00 to 2:00 if you have any questions about the material. I received my PhD from NYU, stayed a bit longer in New York City, got my first job at Columbia before coming here to MIT.

I'm interested in a variety of topics and different kinds of markets, so predominantly in what I call decentralized markets. So these are markets where participants face some sort of friction. So, for instance, consumers may not know what prices in the market are, or consumers and producers have to find each other in order to trade. This has, for instance, applications in transportation markets, on which I have done quite some work. Also very interested in digital markets where, as I'm sure if you have read the news, you've seen that there are a lot of interesting regulatory topics, topics concerning antitrust but also thinking about today's lecture, a lot of challenges, how we actually empirically address these kinds of markets. So I think there's a lot of interesting work to be done there. And then I'm interested also in consumer financial markets.

So that's it about me, OK? So the goal of this and the next lecture is to teach you the essentials of static demand models. Hopefully, you get some appetite for this. And then you come back to [14.]273, and you learn a lot more about this. And I will really focus on the kinds of models that people use instead of giving you a long history of what people have tried beforehand that didn't quite work out. I'll just contrast on a few slides how the thinking has evolved. But I'll be very brief on this. OK, so why model demand? I've already mentioned we might, for instance, be interested in consumer welfare analysis. So any normative question or many normative questions need some notion of demand if we want to know what happens to consumers.

Another reason is that we may want to back out the supply parameters, such as marginal cost or with more complicated models that we're going to learn about in [14.]273, fixed costs, OK? The problem is that this kind of cost data, even if you were looking at publicly-traded firms, is oftentimes not available. If we think about a firm, let's say, introducing a new product, they're not just taking into account the material costs and the production costs of that product. They're also going to take into account the opportunity cost of, let's say, cannibalization of other products that they are competing with. And this would not be captured in accounting data.

You may have heard about a recent literature that has claimed that the US economy-- that markups are rising, which could be indicative of increasing market power in the economy. And so there's a paper that has made a big splash by Jan De Loecker and Jan Eeckhout, which basically shows these rising markups, and they take the production approach. That means that they're basically starting with a production function.

And now there's some more recent work that actually goes market by market and studies, let's say, cars about 30, 40 years and asks, how do consumers fare over several years in the US car market? Is the US car market becoming less competitive? Counterfactual analysis, I've already talked about this. And sometimes you also want to predict demand for new goods with the kinds of models that we're going to study, you can actually do.

So before I jump in, what I would like to do, and I think this is helpful, make you understand what is the goal of these models? So what are these models aiming for? And there's a big set of objectives here, and these objectives are oftentimes actually somewhat in conflict with each other, OK?

So first, the models are to handle a lot of products. We don't want to be restricted in the number of products that we can actually handle, OK? Also, we want to estimate them with aggregate data. So what this means is we cannot see a specific consumer making a purchase decision, and we cannot link specific consumer demographics or other situational factors to a consumer. So instead, we have data of the sorts of what cars did people purchase in Massachusetts in 2021 over the course of a year, let's say.

So the fact that we only have aggregate data is actually in conflict with another goal, which is that we want to generate realistic substitution patterns. So if I observe specific things about consumers, for instance, what their income is or where they live, then I have a much better idea of how they would be substituting. For instance, let's say, I think about stores around Kendall, Aceituna, let's say. And I know that you live around Kendall. I would know that you're probably more likely going to respond to a change in price of a store here than somebody else who lives, let's say, at Central or Harvard Square.

But with aggregate data, we don't have that information. We don't know what your income is. We don't know where you live. And so it's harder to really pin down how people are going to substitute. And as we will see, this will lead to, for the models that people have been estimating, before moving to these modern models to unrealistic substitution patterns because we don't have all the detailed information. And we're not using it in ways to generate those substitution patterns.

We also want to allow for unobserved product characteristics or sometimes referred to somewhat confusingly as demand shocks. So these are basically things that the consumers observe, that the firms observe, that we do not observe as the analysts or econometricians, OK?

So, for instance, if you think about the market for smartphones, Apple will take into account that people like the Apple ecosystem, and therefore is able to raise prices. And so these are the kinds of things that are important to have in the model. But this causes an endogeneity problem that I will talk about next. We also want to maintain computational tractability, which is, again, somewhat in conflict with the demand to have many products and also with these realistic substitution patterns because, as it turns out, that makes the models very nonlinear and harder to deal with.

So just to give you an idea, what was the goal when people constructed these models? These are basically what people try to accomplish. This has been tremendously successful. You can go through the literature, and it's not just narrowly in IO. You can go through pretty much any field and any kind of question that you're interested in, and you'll find people estimating these models to answer questions of all kinds. So from transportation to more IO topics, market power mergers, network effects, to things like residential sorting, and even voting, OK? So now I see more and more people are using exactly the kinds of models that we use in IO to study questions in labor about monopsony power.

So very briefly, front and center-- and I'm sure you have encountered this somewhere-- when we estimate demand, we have to deal with the endogeneity or simultaneity problem. So what does this mean? So there's an interesting article that is actually fun to read because it really shows you how much our thinking has evolved and how much better understanding of even the most simple model that we can come up with, which is this linear supply and demand curves, how we can think about the kinds of empirical challenges that come up here. So start with a bunch of points. You want to estimate the demand curve.

So let's say we have a simple linear demand and supply curve. We have some fixed intercept. We have some coefficients, some slope coefficients. And we have an unobservable part of the intercept, OK? So let's say I'm running this regression. What problem am I running into? Just estimate this via OLS. Conceptually, what is the problem?

STUDENT: They're jointly determined in equilibrium.

TOBIAS SALZ: They're jointly determined in equilibrium. That's right. So this is basically the economic reason why there is a problem. And what this means econometrically, so because they are jointly determined, the price is going to reflect information that's unobserved to the researcher, unobserved information that's in this random intercept.

So if I write down the equilibrium price, I get an expression that contains both the random part of the demand and the supply intercept. And so what this means is that the price is not uncorrelated to the error term of this regression.

How do I fix it?

STUDENT: Instruments.

TOBIAS SALZ: And what would be a good instrument here?

STUDENT: Supply and demand shifters.

TOBIAS SALZ: Right, so we want to find a shifter on the other market side that we believe is orthogonal to the error term, the supply shifter that is orthogonal to the error term. So you may be wondering why am I bringing this up here? You have probably encountered this beforehand, probably even during your bachelor in undergraduate.

So the reason I'm bringing it up is that you'll be fixing this problem over and over and over again. So we will go to much more complicated models than this one here. And it turns out that you can't really escape it. I sometimes hear in talks people say, oh, OK, so now, I've sort of exhausted my exogenous variation in the data. Let me now put a model on top of this and fix the problem this way.

It's almost the opposite of what's true. You actually oftentimes need more rather than less instruments for the kinds of models that we estimate. So we will think very carefully about what instruments you need, how many instruments you need, what are good instruments, and things like that. So just to prepare you for-- you will not be able to somehow get around this problem with the kinds of models that I'm going to talk about today.

So we're going to estimate demand systems. And the initial approach here was what is called the product space approach. What is the product space approach? The product space approach-- and I'll just have one slide on this. I'm probably not doing this literature justice, I should say. Just very quickly want to give you the gist of what people have been doing here, OK?

The product space approach basically means that I treat every product as its own unique thing. And I'm trying to model a demand system where I write the quantities demanded in terms of all the prices in the market, in terms of all the demand shocks in the market and all the unobserved shocks in the market.

So let's think about already a much more restricted version than this very general formulation here, one where, let's say, the vector of quantities for different products in this demand system is a function of some matrix that premultiplies a vector of prices. So we have a market with many different products, let's say the car market since this is going to be our theme today. And so we basically, to understand the demand for each separate product in this market, I need to know the coefficients and all the prices for all the products.

And so then there's going to be a scalar unobservable also. So this is already, again, a restriction because we could-- in principle, it may be true that the demand for each product depends on the whole vector of unobservables. So what is the issue, even in this very restricted version here, is I have a lot of parameters that are, again, unique. It's a unique set of parameters for each product. I need to estimate all of those.

And so if I increase the number of products, I very quickly run into a curse of dimensionality because the number of parameters grows at a squared rate. So each new product that I'm adding to the market, I need to add in a row and a column to this matrix.

So people have tried to reduce this problem, for instance, by restricting the matrix such that the substitution patterns are consistent with Slutsky's symmetry, or they introduced ideas like budgeting. So where people, let's say, have some budget for food and some budget for leisure and some budget for something else. And then the demand in each of those categories, you only need to know what's the budget for this category. And that restricts substitution patterns.

But by and large, you don't get around this curse of dimensionality. And it has therefore turned out not to be a very practical approach. It's also very demanding in terms of instruments, if you think about it because all the prices here are potentially endogenous. And so for each product, I need the number of instruments equal the number of prices, OK?

Another reason why people don't like these models-- I find this as somewhat less good justification because it's not that we do it all that much, that we can't predict demand for new goods because, again, I don't learn anything about other products by learning demand for the products that I'm currently estimating because there's nothing that translates.

OK, so people have moved away from this-- although, there are nice applications of these kinds of models-- and instead move to something that is called the characteristic space approach, which typically comes together with discrete choice models.

OK, so in this approach, products are described by their characteristics. So take the market for cars, let's say. So the car can be described by horsepower, top speed, trunk space, driving assistance features, and things like that, OK? The iPhone is a combination of, let's say, the megapixels of your camera, the resolution of your screen, your storage, and other kinds of gadgets that are built in. So we enumerate different product dimensions. Now, you will notice that if we add products to the market and we observe what are the product characteristics, we in principle are not adding any new parameters.

So this idea of having products basically be collection of characteristics, as I said, comes together with what is called the discrete choice approach, which is that people-- there are a bunch of consumers. Each of them buys either one of the available products or none at all.

So to summarize then, we can write-- sorry, I should have mentioned, a third foundational principle is that we want to build up these demand models from utilities, from indirect utilities. So we're describing people's utilities in terms of prices, how much they value each of these different product attributes-- so these are observed product attributes-- the unobserved product attributes. These are in IO religiously referred to as xj, OK?

So everybody knows what xj is. And soon you'll also know that whenever you seek xj, these are unobserved product attributes, and demographic variables as well as unobserved random shifters at the consumer level. And then we make the assumption that consumers-- in most models, you make the assumption that consumers do the best they can and pick the product that they like the most.

OK, so I was-- let me just go back once more here. So there's this sort of random part, the consumer-specific random part, that helps us explain why consumers may choose different kinds of products. It basically allows us to rationalize different patterns in the data. But it's random from the perspective of the researcher. The consumers, they know what their idiosyncratic valuations are for each product. So I'm going to just switch notation here just for this one slide, because I couldn't find an image like this with epsilons.

So suppose that consumer utility can be written as the random part, which is mu ij. So each consumer goes to the market, takes a random draw for different kinds of products. And they all face the same price. So what you then can do is you can describe the consumer's choices in terms of these regions, which then give-- these are essentially measures that give rise to the market shares that we observe.

So, for instance, there are some consumers who drew both a low mu i1 and a low mu i2, and they don't purchase anything because in both cases it's below the respective price. There are some other consumers who drew a very high mu i1 and a smaller mu i2. And they're in this ai1 region here.

So we basically get these-- in terms of this distribution of the random part of utility, we get these choice regions that tell us what market shares are. So one reason I'm showing you this here is that if you want to compute market shares or choice probabilities, you have to come up with a way of finding these regions, in principle. You would have to figure out, in this region of integration of these different epsilon draws this is what consumers choose. And this other region is-- they choose something else. And you, in principle, unrestricted functions for these epsilon, that's a very cumbersome exercise and, in fact, computationally very difficult to do.

OK, people have solved this the following way. So coming back to our utility function, I'm now just isolating the random term. What was beforehand called mu ij is now called epsilon ij. And it's typically assumed that this error is type I extreme value. So you may now be wondering what product-- what problem in life could possibly be solved by a double exponential functional form?

As it turns out, this has a really nice property, which is that if we assume that consumers do the best for themselves, they observe everything, they choose the product that they like the most, then the choice probabilities or the market shares can be expressed in this logit form, OK? So we get a closed-form solution for those probabilities.

Now, we can make some normalizations here. If you stare at this, you'll notice that we could add a constant to each of the uij's, and this constant would cancel out here, which means that we have a normalization to make. And typically, we normalize what is called the outside good to 0.

So the outside good is basically everything else that consumers could be doing instead of buying some of the products. And so that's important. It's not just a mathematical decision to normalize this to 0. You have to always think about that consumer welfare is measured relative to this product. And it's therefore worthwhile to, in the interpretation of a consumer welfare calculation, keep that in mind. And that could sometimes be a problem if you compare markets from year to year because the outside option could actually change.

So maybe as a question to you, let's think about the market for cars. So what could be problematic about this? So what is the outside option when we think about people purchasing cars? What is part of the outside option? Yeah?

STUDENT: Public transportation?

TOBIAS SALZ: Public transportation.

STUDENT: Cycling. Cycling.

TOBIAS SALZ: Cycling. I'm looking for something more long run. But all these are good responses. So most people own a car already. And the car is a durable good. And the decision whether or not to buy a car is in part a decision of how much I let my old car depreciate. And to what extent people are willing to do that might vary with, let's say, macroeconomic conditions, how easy it is to get credit, things like that.

So if I now estimate over several-- if I estimate a demand model that spans several years, where this decision of consumers may change from year to year, I have to keep this in mind. I may have to do some adjustments for a substitution to the outside option from year to year. So that's something that's important to keep in mind.

So what is the typical data scenario for these kinds of models? As I said, researchers observed aggregate market shares, prices and product attributes. For each market, they also observe or tend to use the distributions of certain consumer characteristics. These characteristics are not linked to a specific purchase decision. So I may know that the income distribution in Cambridge looks like this and that the family size distribution in Cambridge has this form. So these are things that you typically can get from census data or somewhere else.

I'll talk more about instruments later. You need instruments. Importantly, because we're working with aggregate market shares, the typical assumption, although there is some recent work that tries to relax this, is that we can ignore sampling uncertainty in market shares. So we're saying we observe enough people buying each product that I can just take this share as given and do not have to econometrically deal with the fact there's some sort of sampling uncertainty. We just take these as numbers as given, OK?

Typically, this requires some judgment. You have to make a decision on what is the size of the market. So, coming back to my example, in the market for cars, it's not immediately clear from year to year who's in the market to buy a car because it may depend on how much longer do people want to keep their old car and a variety of factors like that. So this oftentimes requires careful judgment of what is the size of the market. In some cases, it's a bit easier.

For instance, let's say you have access to scanner data. You see how many people go to the supermarket. And you would be interested in, let's say, the market for milk. Then you know how many people have been in the store, and it gives you a pretty good idea how many people could have bought at this point in time. But you have to make assumptions on that because counterfactually you may be interested in something that either lowers or raises prices. And then how many people substitute from the outside option into the market and purchase a product depends on the size of the outside market.

So what I'm going to do now is I'm going to show you the model that people estimate these days, OK, the canonical model. And then I'm of trying to go through the different pieces and tell you why do we need those different pieces? So this is the random coefficient demand model. It goes back to-- or it has been popularized by a paper by Berry, Levinsohn, and Pakes, much of the intellectual contribution of how to estimate these models and deal with some of the econometric challenges has have been sort of laid in this paper.

So, as before, we have an indirect utility function, which is assumed to be linear. We have consumers i in market t purchasing products j. There again are product attributes, prices. And notice that these are now premultiplied by a coefficient that has an i subscript. So what this means is that this is a model that allows consumers to experience different utility from different product attributes, OK? There's some people who-- supposedly, those with families, they're more interested in having a lot of trunk space than some other people, OK? There are some people, supposedly, those with lower income, that are more price sensitive.

So this is one of the key innovations of this model is that it allows different people to assign different value to both product attributes and prices. And then this introduces a lot of challenges. And then how do we deal with those challenges? So we have, again, the outside options normalized to 0. So the utility for the outside option is just given by the random part, the idiosyncratic random part. The epsilons R seem to be type I extreme value. x is a row vector of observed product characteristics. And then xj, xj is a product market-specific demand shock.

And so these coefficients, they are typically assumed to be a combination of an intercept, the average, let's say, the average price sensitivity plus some things that scale with the demographics. Again, I already told you we might observe the income distribution. And so the important thing, though, is that we do not observe income at the consumer level. We just have the distribution, and we integrate it out, OK? So this coefficient depends on demographic variables that we observe, again, at the market level and then, in some cases, also a random part.

So typically, this is referred to as nu. And most of the times, this is assumed to be either normally distributed or log normally distributed. And then similarly for price. Any questions up to this point?

STUDENT: So tastes and price sensitivity only vary across markets?

TOBIAS SALZ: No, no, no, no. They're buried within markets.

STUDENT: Within markets.

TOBIAS SALZ: Yeah, because, again, we have different consumers in each market, indexed by i. Again, we don't know that this purchase was linked to this consumer with this income. But I do know that this is the income distribution around Kendall Square. And I therefore know that some people in this market have this much income and so on. And I'm just integrating it out. Other questions?

STUDENT: So the superscript k is just--

TOBIAS SALZ: This keeps--

STUDENT: --indexing on product characteristics.

TOBIAS SALZ: Yep.

STUDENT: OK.

TOBIAS SALZ: OK. OK, so this is the model. Now, let's drill down a bit and understand why do we need these different pieces? So first of all, it's going to be useful to keep track of the things that are i-specific and not i-specific, so things that vary at the consumer level and things that are sometimes referred to as just the average product quality.

OK, the average product quality is often to refer-- pretty much always refer to as delta. So one of these things, as xj people, when you say delta, they know what you're referring to in the model. So this is the average product quality. It includes basically the intercept of the random coefficient and the unobserved product characteristics. And then, in this case, we keep track of the consumer-specific part as mu ijt. And it's, again, multiplying both attributes and prices by these consumer attributes that are either unobserved or observed.

OK, so now we can take a specific consumer. And, as before, this consumers purchase propensity for product j is simply given by this logit formula. But they're, as I said, different consumers within a given market, t, right? And so the overall market shares have to integrate out of all the different consumers. So we therefore integrate out over the distribution of demographics as well as if we allow for unobserved distributions over those as well. OK, so this thing here is much more nasty than this thing, as it will turn out, OK? I'll show you in a second why that is the case.

STUDENT: So we set the distribution for nu.

TOBIAS SALZ: You make a functional form assumption, but you don't know what the variance of that term is, yeah. OK, so question, why do we need this error term xj? If you think about it, we know what these-- we know how we compute probabilities.

And if we only had observed product attributes and prices, what would happen is that our model is very quickly rejected by the data because what it would imply, because again, we assume we have many consumers in the market. We can therefore just write choice probabilities in this logit form. And that means that I only need to know x, beta, alpha, and p to know what the market share of a product is going to be.

But what this implies is that any market where two products have the same attributes and prices, they would need to have exactly the same market share. Or, to pick a different example, if I look at a market where all the product attributes of one product are worse and that product has a higher price, that would not be allowed to happen in a model without such an error term.

So think again back to the iPhone. So Apple, in the past at least, tended to not introduce new features as quickly as some other producers. So the camera didn't have the highest megapixels. And it didn't use the biggest storage space, and it wasn't the largest phone, and things like that. And yet, Apple had higher prices and larger market shares, at least in the US, than, let's say, cheaper Android phones.

So without such an error, which is a purely vertical attribute of a product, that all consumers value the same, something like this couldn't happen. So we need xj in order to explain that there are some products that somehow are popular, even though they don't quite look very popular in terms of product attributes.

This, of course, causes this problem that I've been talking about earlier. Firms-- Apple also knows it has this ecosystem. It has a good product and is therefore able to raise its price. And so this creates the endogeneity problem. Any questions about this.

So why do we need the random coefficient? Again, this is one of the big, intellectual pushes in this paper is to introduce the fact that different consumers are allowed to value different products as well as prices differently. So think for a second about a model in which the coefficients do not have an i subscript, OK?

Now we are allowed to write the market shares, not just the individual consumer probability of purchasing a product. We can write total market shares with our logit formula, and it's simply given by these delta terms, which is the average product quality, remember.

It turns out, if we have aggregate market shares and is logit assumption, we get something really nice because we can just take the log of each of the market shares and subtract from that the log of the outside market share. Remember, we normalized the utility of the outside market share to 0, OK? So log of exponential of 0 is 0. This term here cancels out because it's in both the market share for the inside goods and the outside good, OK? So the denominator here cancels out.

And what we get is on the left-hand side, log market share for each product minus the outside log of the outside good has this linear form here. So why is this so nice? Well, this is something we observe. We observe the aggregate market shares. I can just call this y, right? This is just a linear regression.

So in a simple logit model, again, with aggregate market shares, we can just transform this whole thing into a linear equation. And we can actually, if we believe that prices are exogenous, which typically they're not, you can estimate this with OLS or otherwise with two-stage least squares. And we would be done, OK?

So the simple logit model without the random coefficient has this great feature, that it would make things very simple, very transparent, much less room for coding errors, OK? So why do we go through all the lengths of avoiding this model? So I'll try to explain this first just in terms of some of the proper mathematical properties of this model. And then I'll also give you the intuition.

So the problem is that the simple logit model generates unrealistic substitution patterns. So, for instance, if we're interested in price sensitivities, so how much does product j respond to change in the price of product k? Then, with a simple logit formula, this is simply given by alpha naught, the intercept of the price coefficient, right, or, in this case, the price coefficient and the product of the market shares of the two products, k and j.

So why is this problematic? Actually, you get something similar for the diversion ratios. That is, how many people substitute from j to k if I change the price of product k? This is, again, only a function of the market share of product j and the market share of product k. Why do we not like this? So think, for instance, about a market where you have a spectrum of cars. And you order them from low quality to high quality. It may just turn out that the market share of the crappiest car is 10% and the market share of the best car in the market is also 10%.

And now we're asking, how do people substitute from these two different cars? The Yugo-- this is a famously crappy car described in BLP. And let's say at the top end you have something like a Lexus or a BMW. Both of them have 10% market share. And you're asking, how do people respond to a change in prices of some third product, some third car, OK?

Let's say you take the Lexus and the Yugo, and you change the price of a BMW. We would think that people who would purchase the Lexus and the people who purchased the Yugo would respond very differently to a change in the price of a BMW. This model does not allow for that. So it's not just about all we want the-- we want the model that's a bit more flexible. This is a really substantive restriction, that, for the kinds of questions that we're interested in, which is what happens if Lexus and Audi or Lexus and BMW merge, is crucial.

So the reason we pick this functional form is really just for analytic tractability. There was no economic reason to choose the logit error terms. This was to avoid having to keep track of all these complicated regions of integration. We have to get a nice analytic formula for the choice probabilities. That was the rationale for picking this functional form.

So basically, we want to fix this. So let me give you one more intuition for why this happens in the logit model, so going back here once more. So what leads consumers to make different choices in this model? Why do different consumers in this model make different choices?

STUDENT: The epsilon.

TOBIAS SALZ: Only the epsilon term, right? That's the only reason why somebody may pick a Lexus over a over a Yugo. But the epsilon terms are assumed to be i.i.d., OK? So the fact that you purchase a Lexus, it doesn't reveal anything about you because next time you go to the market you take a new epsilon draw.

But what we want is, of course, we believe that somebody who buys a Lexus, there's something about that person that is different from a person that buys a Yugo. And the model does not allow for that. So that's why we need these random coefficients because now we do allow for the possibility that somebody with higher income is less price sensitive. And since they have a lot of income, they might as well get a car with more fancy features, better safety, and things like that, OK?

Now, as it turns out, things get a lot more complicated. The formula for elasticities, which was just to show you this again, previously, again, just a function of prices, market shares, and the price coefficient. We now have to integrate out over all the different consumers and have to keep track of their purchase probabilities for our product j and k if we're interested in cross price elasticities or product j if we're interested in on price elasticities.

So different consumers make different choices. And for a given guess of parameters, we get a distribution of nonlinear of these nus. And then, this leads to a weighted version of these elasticities, where different people have different substitution patterns.

So let me talk about consumer welfare. Similarly, as the choice probabilities, if we took a random distribution, this would be a very complicated object because we have to keep track of-- consumers go to the market. They draw some of these epsilon terms. They then make optimal choices. And they have to form the expectation of that. Again, I have these regions of integration that I would have to keep track of and what consumers choose.

Again, the logit model drastically simplifies this for us because we get something that's called the inclusive value formula, which tells us what the expected welfare is before consumers take these epsilon draws. So in expectation, get some draws, make optimal choices. This is the formula that you get for welfare, OK?

Now, this is measured in utils. You can convert this into dollars by dividing through by the price coefficient. There's something that's-- this introduces some practical problem. Because of the-- and sorry, one thing I forgot to mention, this is just for one consumer i, OK? So in the random coefficient model, we would then have to integrate this again out over the distribution-- of those distributions of nus and ds.

So there's something that people have found to be problematic in the literature, which is that if you have counterfactuals that drastically change the number of products, then you get, oftentimes, strong welfare implications from just adding and subtracting products because people harvest these epsilon draws, right? And the problem is that these epsilon draws are both an econometric error term that helps us rationalize choices; but also, we give them an economic interpretation.

This actually, again, something that you have to, in your work, really-- it takes some getting used to to think that-- to understand that we actually give an interpretation to the error term. And it's part of the model. It's not just tacked on. And in this case, we typically assume that consumers know these epsilons, and they're not just errors. They actually take this into account when they make choices. And so this means that adding more products, right, that can harvest more of these epsilons, and it increases welfare.

I don't want to make too big of a deal of this, but it's just good to be aware of this problem. Some people argue that the random coefficient model actually ameliorates this problem because, again, people, they're somewhere in a specific part of the product space by virtue of having a specific coefficient determined by income and other things. And they're not going to just substitute globally. But they may be substituting more locally. And therefore, the epsilon harvesting is not as big of a deal.

So earlier I said that we want to think carefully about instruments. And again, what's the reason? I gave you this Apple example. Apple knows there's some secret sauce that makes its products better than some other products, and it might price this. So both the researchers-- sorry, the consumers and the firms observe this xj, the researchers do not.

So this then raises the question, what makes good instruments for these kinds of models? And how many do we actually need? And you may be surprised by the second question because sometimes one is not enough. So if we just take our simple logit model, where we do not have random coefficients, we can do this inversion that I talked about. It gives us this linear form. And you stare at this, you say, well, this is just a linear equation, a regression equation. And you're right. We can estimate this if only prices are endogenous with one excluded instrument. So in this model, if you have a supply shifter that you think is uncorrelated to xj, you can just estimate this model, OK?

The same is, unfortunately, not true for BLP, in general. We will go in [14.]273 and dive deeper in the identification of these models. So today I'm just giving you the intuition. So there's two ways to provide this intuition. One is, well, with nonlinear parameters, I can't quite separate things like this. And the demand for each product depends on the whole vector of unobserved product attributes. But perhaps a bit more intuitively, if I just have a supply shifter and change price, that does not necessarily tell me about whether somebody who purchases a Lexus substitutes to a Mercedes instead.

And so for the nonlinear parameters, we want to identify the variants and the interaction terms with income. We want to know in which part of the product space do people substitute? And a simple supply shifter that's the same for all firms wouldn't do that, wouldn't tell us about that. So we need something that tells us something about where are these consumers that purchased this product would be going to alternatively, to understand-- to identify the nonlinear parameters. Any questions so far?

OK. OK, so let's talk about some different instruments. So typically you get very good variation from looking at different markets, where in some markets, let's say there is a Lexus dealership and in some other markets there is not because then I can see in the markets where there's no Lexus dealership the market shares of which other products is changing the most.

So that's usually very good variation. Can you think of a concern with this argument? Yeah.

STUDENT: Could you expect the presence or not of the Lexus dealership to be correlated to [INAUDIBLE]?

TOBIAS SALZ: Exactly, right? So we think that, at least in the long run, the Lexus dealerships open where people have high demand for those types of cars. And actually, people have estimated models where they look at combined, the antiproblem and the demand problem combined to make use of exactly that idea. That's precisely the problem. We think that firms know which markets are good ones to enter. And so this is something that's probably more likely to hold in the short run.

Another set of instruments are cost shifters, so material costs, distribution costs, tariffs, taxes, things like that. These are oftentimes very good instruments. They have a clear story for relevance and the exclusion restriction. So the problem is that they're oftentimes not available. It's hard to learn about cost, as I actually argued earlier.

So what sometimes people do is they argue for the following approach here, which is, OK, we may not observe cost, but maybe observe a proxy for cost, OK? These are called the Hausman instruments. So the idea here of these instruments is that if I have a cross-section of markets, I can look at a specific firm and can look at the prices of that firm in other markets, not the market of interest. So I take the average of prices in other markets, and I can take this as a proxy for cost shocks for this firm.

So maybe in this specific year, the supplier of this manufacturer, I don't know, went bankrupt, and that caused some trouble and therefore have higher costs. And so that's reflected in all the prices of this firm across different markets. I do not observe this directly, but the average prices in other markets, in some sense, part of my data. Can you think of concerns with this reasoning? Yes.

STUDENT: Prices are also based on demand for each of them?

TOBIAS SALZ: Exactly, we're back to this problem that we had initially, right? Prices reflect both demand and supply factors. And this line of argument only works if the costs are very correlated across markets and the demand shocks are not. But then these Hausman instruments are convincing. Otherwise, you have to think hard about whether in your setting this is a good idea.

Then there are what I call the BLP instruments. Those are the instruments that are actually used in BLP. So these are the attributes of other products in the market or some function of those attributes, so either the sum or the average or-- there are various ways of constructing them, in fact.

So the idea here is all along we have been assuming that only prices are endogenous and product attributes are fixed. So now, if that's true, again-- you have to think about whether that's a good assumption-- then I can use the fact that I'm competing in a market where I'm trying to sell a very fast car to consumers. But there are also a lot of other producers that also offer fast cars, that the product attributes for other cars are excluded from the utility function of consumers that want to purchase my product. But they force me to respond to the fact that there are a lot of other fast cars in the market.

So that's the idea, that it tells me how, in terms of the things that consumer value, how much competition I'm facing. And so I've sort of already alluded to this. The concern is, of course, that the product attributes are not in fact exogenous. So coming once more back to my favorite example of Apple, Apple has, again, as I said earlier, typically delayed the introduction of new features. And it knew it can do that because they're other things, unobserved things that people like about Apple, right? And so, again, you have to ask yourself, is it actually a good idea that product attributes are exogenous?

There is a whole literature on how to construct these instruments. Again, this is the economic idea in a nutshell. There are ways-- some people have argued just taking the average or to leave one out is not a good idea because if you look at large markets, to leave one out is actually the same for all firms or close to the same. And then you want to basically take differences instead. There are also questions about how do you-- if we are in a many weak instrument situation, how do you construct the most powerful instruments? Or how do we approximate them? There's a paper by Gandhi and Houde on doing that. Again, in [14.]273, we'll talk a bit more about this.

The last instrument that people sometimes use are what I call the Waldfogel or Waldfogel- Fan instruments. This is the idea that firms don't set prices just for one market, but they have larger zones for which they set prices. So for instance, a dealership network that spans both Massachusetts and New Hampshire may set prices for its cars both for Massachusetts and New Hampshire. Although, these are somewhat separate markets. You can argue with that as well, of course.

And then the idea is that there are a lot of people in New Hampshire who like pickup trucks, and there are fewer people in Massachusetts who like pickup trucks. But the fact that there's a lot of demand in New Hampshire, combined with the idea of zone pricing, leads to higher prices for people in Massachusetts for pickup trucks. Although, that doesn't necessarily mean that they like pickup trucks. So that's kind of the idea.

Again, there's a nice paper that actually documents that firms do price very coarsely across markets, do not take into account local demand variation as much as they should in a profit-maximizing way. There's an obvious concern, of course, with this instrument, that those demographic factors and what drives demand across markets is also correlated, that it's not as separable as we think.

So I'll use the last 10 minutes for today to talk a little bit, how do you actually do this? I told you everything about instruments, but I actually haven't told you how you use them in this model. You know how you use them in two-stage least squares. But how do we use them in this complicated model? So remember, if we can-- in a simple logit model, we can do this inversion and express everything in a linear equation.

So here, the problem is more complicated. So BLP is typically estimated via GMM. Although, there are other ways to estimate it that I'll, again, talk about in [14.]273. So this is a GMM problem, where I minimize some vector product of moments that may be weighted by some matrix omega. And those moments are simply-- these are the empirical counterparts to the moments. I'm interacting my instruments with the error term.

Now, what is the error term? I can use-- remember, the definition of delta was it's all the non-i stuff. It's the average valuation of consumers for a product j. So I can subtract from that the product of attributes and coefficients that are non-i-specific. And I get back my error term. Remember the definition from earlier. So that's my error term.

OK, but that's not super helpful because delta is stuck in here. So if I want to interact my arrow with the instruments, I need to know what delta is in order to know what delta is. And it's kind of stuck in this integral. Or the empirical counterpart to this would have a sum that approximates this integral.

OK, so what do we do? I'm moving this to the side here. It's the same stuff but a bit more compactly written, OK? So here's how you estimate BLP. Again, there are many details that are left out here that, if you want to actually do this, you need to learn about. So the outer loop is just minimizing this thing here. And so, for reasons that I'll come back to in a second, in the outer loop, you only guess what I call the nonlinear parameters. These are the parameters not counting beta naught and alpha naught. These are the parameters that tell us something about how much consumers substitute across different products, different consumers differently across different products.

So I'm guessing some nonlinear parameters. I fix a guess. And then, in the inner loop, I'm solving the following problem. For every market t, I want to know what my deltas are because if I know what my deltas are, I can build my size. If I know the size, I can build my moment conditions, OK?

So what are the deltas? So this is where this paper, among many other things, does something really clever. They show that it can start with any initial guess of those deltas, OK? Start with any guess. And then I can iterate on the following thing. Start with this guess. Compute the market shares, the predicted market shares from the model.

And then add to my initial guess of delta the difference between the actual market shares that I observe in the data and the predicted market shares. This gives me a new delta. And then I do this again and again and again, and it converges. In fact, this is a contraction mapping. So for any set of nonlinear parameters, it converges to a unique thing, that rationalizes the observed market shares exactly under this guess of parameters.

This does not tell me that this guess of parameters is actually the one that minimizes the objective function. So in the outer loop, I still have to guess new parameters until the objective function is minimized. So you keep on doing that. The one thing I said, well, you only have to guess the nonlinear parameters. So one trick that people use is that if you know the nonlinear parameters and you do the contraction mapping, then it's as if I know delta, at least for this set of parameters. And then I can just run a regression here because for any given theta 2, delta is a known object. I can regress it on this. I don't have to search over those linear parameters here. This oftentimes leads-- you can speed up this problem quite a bit, OK?

So you iterate on the outer loop. You guess new parameters. You iterate on the inner loop. You find your deltas. You run the regression. You built the moments. And you keep doing this. Any questions on this? This was fast. You may understand it at a high level. I encourage you highly to actually do it. Come to [14.]273, and you'll learn how to do it.

STUDENT: I guess I'm curious when this would fail.

TOBIAS SALZ: There are many ways that it could fail.

[LAUGHTER]

One of the most memorable things in grad school is when we all coded up BLP together, so the first time. Many things that can go wrong.