

MITOCW | 12. From Reduced Form to Structural Evaluation

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

AUDIENCE: Was that going to be-- were we going to do that session after you do this?

ROBERT No.

TOWNSEND:

AUDIENCE: [? Or ?] [? will ?] [? we ?] do it before?

ROBERT We'll get your questions first. That way, I can't time how quickly or how slowly to go through this lecture. So I'd

TOWNSEND: just as soon start off.

AUDIENCE: I have a question.

ROBERT Yes.

TOWNSEND:

AUDIENCE: I wondered if you could explain in more depth how-- so, when you were-- when you're simulating structural models, and there's an obvious distinction between structural models in which people are forward-looking and which people are not. And it makes it easier if people aren't, but it's not as realistic.

So I wondered if you can kind of explain like what are the trade-offs in making people forward-looking and then how you-- I think there was like a [? Vieira ?] paper about how he was able, for the first time, to like get at the transition and not just steady state. So I didn't understand [INAUDIBLE].

ROBERT Yeah, OK, so there's like three or four different things in there. It's a good question, questions. First of all, this

TOWNSEND: forward-looking or not forward-looking, part of it is just tractability. The first generation of these models, half of them basically assumed what's called warm-glow bequest motive. So you just sort of feel warmly about passing your money on to your kids, but, actually, it's you, and you are reborn the next period.

We went through LEB, which is this Lloyd-Ellis and Bernhardt, which is a simple sort of bequest motive. So you don't take into account the utility that's generated by the bequest. You just model it as a reduced form. It's a bit like overlapping generations, and that's why overlapping generations models are still widely used.

We had another. There was a *QJE* paper on financial volatility, which was overlapping generations, another one on China and savings rates, which was overlapping generations. It's just far easier to solve a two- or a three-period problem and splice together these overlapping generations than it is to have a forward-looking behavior. So it's a judgement call.

Now Paco, way back to his dissertation, but also then subsequent papers with Kaboski and so on, have this forward-looking aspect to it. So not only do you have to kind of work backwards on your own decision problem, but, to the extent that prices and wages are endogenous, you kind of have to forecast those too. And, in rational expectations, you would require that the forecast be rational, that households are kind of getting it right on average.

Actually, that reminds me to say that one of those flow of funds papers we reviewed briefly was Monika Piazzesi and Martin Schneider. And, although they were using US data, they were looking at the effect of various things, including a demographic shift, sort of baby boom generation having more young people in the population, relatively speaking, than before. And they did not have the forward-looking behavior because they basically assumed that the expectations you see in the Michigan surveys are accurate. So--

AUDIENCE: What were the expectations of?

ROBERT Prices.

TOWNSEND:

AUDIENCE: Would they go up?

ROBERT Inflation rates. So it wasn't like solving for an endogenous price in the model and then having people forecast

TOWNSEND: that. They just short circuited the whole algorithm by assuming that people's expectations in the models were as measured in the data, even if those decisions that got generated might be inconsistent with those inflation forecasts.

And it was a reminder to many of us how straight-jacketed we've become, in a good and a bad way, with the rational expectations because, once you have a way of pinning down the future via these measured expectations, you can expand the heterogeneity in the model and deal with a lot more realism.

AUDIENCE: So, in that inflation, how far did they forecast inflation in the survey?

ROBERT I think it was just the following year. So the last thing you asked about were the transitions.

TOWNSEND:

AUDIENCE: [INAUDIBLE]

ROBERT Sorry, I didn't hear you.

TOWNSEND:

AUDIENCE: Oh, I was just telling him they measured for different ranges of times. I was like one year ahead and five years ahead.

ROBERT They were concerned with consumer durables and housing purchases and so on. So you'd want to have

TOWNSEND: expectations of some of those asset prices.

The transitions are the hardest thing to do. You can have a model, which is forward-looking, and solve for the steady state. It's like taking a little tease off of all the variables. So, you know, things will settle down to one interest rate and one wage rate. That doesn't mean the decision problems are static. People are still forward-looking, but they're looking forward at those prices.

And, in a steady state, there's always these offsetting fractions of the population. So some fraction become urgent. The rest become patient. There's always a fixed fraction. And there's always a fixed demand and supply of funds. And so you get these.

And you can almost analytically, sometimes, get the steady state of some of these models. Or, even if not, you can compute them much more readily, whereas the transitions, by definition, you have a whole path, and you have a path of guessed interest rates and wages and so on. And people take that path as given and optimized.

And then you have to look at the excess demand and supply for labor and for capital not only in the current period, but all these future periods for which they've already made a guess, and then have this algorithm to somehow adjust the guessed path and hope, somehow, it will converge.

So Buera and Shin did that originally and have modified the code. Yan, you've played around with that code. Other people have as well. It seems to work, but it-- if you take it to a new application, you've got to kick it and push it around a bit and get scared that it's not going to work at all. And so it's certainly not a silver bullet, but it does seem to work for particular applications. Am I answering all--

AUDIENCE: Yeah.

ROBERT OK.

TOWNSEND:

AUDIENCE: So I actually have a related question. So I remember when we were talking about the computation issues in one of the extra sessions. We were talking about having myopic agents and saying, well, in the profession, we're not that interested in myopic agents.

But I feel like there is potentially some middle ground between saying, well, I have to have an agent that looks to the infinite future, or I have to have an agent that is totally myopic. And I feel like it's actually not more sensible to say this agent looks to period infinity than to say this agent doesn't look forward at all. So I'm wondering like what way you see like-- is there a reason I'm not understanding why one is more ridiculous than the other? Or like I mean--

ROBERT People take positions. There's no question you're going to get opinions. Some people would dismiss the myopia
TOWNSEND: out of hand as outdated. Other people are convinced there's a lot of irrationality and behavioral things. And I do think that taking the expectations as measured in the data is quite reasonable thing to do.

And some overlapping generations models are, with realistic lifetimes and mortality and all of that, are really, in some sense, the most realistic of all. Granted, you still have to deal with this bequest motive, but, you know, I'm practical. I think it depends on the question at hand. And it's nice to be fully rational and forward-looking and all of that, but computational, dimensional issues will limit the model you want to analyze. And, sometimes, it seems unreasonable.

So I mean, for example, in IO-- I'm not sure if we'll get to this today-- the state spaces just expand dramatically when you start thinking about strategic behavior. And then you can actually show, if you wanted to keep track of where all financial providers are going to locate with a finite number of nodes, kind of working backwards and having-- that becomes a very computationally intensive problem. And it seems reasonable to say, well, let's just have them forward-looking for two, three periods or something, a finite number of periods, as an approximation.

AUDIENCE: I think it's so interesting because, from outside the profession, a lot of the criticism we get is, oh, you assume that these agents are totally forward-looking and perfectly rational. You guys should see what happens in your models if you don't assume that. But then, within the profession, if you were trying to assume that, people would be like this is a waste of my time. Leave me alone.

ROBERT TOWNSEND: Well, the profession comes and goes in terms of its rational expectations, which was kind of invented by Muth in 1960-something I think, late '60s. And then it took over macro with a vengeance, well, in the sense that there were a lot of unhappy campers. So people had stakes in doing certain-- one way or the other.

And then you need to start putting back ingredients in the models to explain rigidities and slow-moving behavior. I mean, personally, I'm not so sure sticky prices is any more compelling to me than sort of some fixed expectation.

So those are-- you'll just have to develop your own judgment somehow, but there isn't a single right way to look at it. I think it really does depend on the applications. And, especially, as I said, if you want to break new ground and explore a different kind of model, you may have to make some approximations. Yes?

AUDIENCE: So I have a more general question, sort of how to interpret one of the classic models that we saw. So it made a lot of sense to me like the last part of the other course where we looked at insurance and how to fit data with different models. But there was that one lecture on TFP that I am kind of not clear exactly how to think about it in the context of this class. So, the paper by Zilibotti and other people on China, how do we-- how should we think about it in the context of this class?

ROBERT TOWNSEND: Well, the way I think about it is we're interested in development and development issues. We're interested in the impact of financial programs. And, for both those reasons, it seems like we have to take into account how the entire economy is evolving.

People don't narrowly stay in their village. They can migrate. They're forward-looking. Potentially, there's urbanization going on and so on.

The starting point-- and I put it at the beginning, rather than the end of the course, was what models do we have. Let's take an inventory of the different kinds of models we have to try to deal with some of the puzzles like why is money leaving China while the rates of return are so high in that paper that you mentioned, "Growing like China."

It is true, for example, that a lot of macro guys are happy sort of modeling TFP as a process. The business cycle literature, following the lead of Prescott and Kydland, is very much like that. The development literature still has those ingredients in the sense of allowing individual talent and productivity to vary, but part of what we see in the aggregate data is an equilibrium phenomenon. TFP has to do with potentially improving allocation of resources across individuals and households.

So that brings us back to the micro data. And the way I see the inter-relationship is I could have equally well started with the micro underpinnings and built up to the macro, rather than starting with the macro and then getting into the micro. It is true, I think, that you need to iterate back and forth.

You tend to lose the perspective when you're just reading macro development papers. So what does this have to do with the poor people in an Indian village? Or, vice versa, when you're trapped in a village, you can't see what's going on in the rest of the economy. I mean, it's just my firm belief that we absolutely have to do both. Although, how much of one versus the other depends on the application.

But this is a macro development class. So it really has to get us all up to speed and comfortable with sort of macro. And the cool thing about macro development is this whole generation of models that's coming along that are explicit about micro underpinnings and potentially testable, whereas so much of the macro literature, real business cycle literature, new Keynesian, and so on, just make assumptions about what's going on at the individual micro level and never test.

Markets, institutions, the nature of financial contracts are all primarily just assumed and taken as given in the more standard macro literature, whereas development is kind of a wonderful playground, if you want to put it that way, for seeing just, whether you're going from urban to rural areas or one country to the next, you can see these constraints being estimated and varying. And the work we did in the class shows that that matters for the macro.

AUDIENCE: Yeah, so I guess that my point of view, maybe also because of my background, but it seemed to me that, the last part, we really were in villages like seeing how people do risk sharing and stuff like that and your paper with Kaboski about the Million Baht program. But then it seemed maybe we didn't do that much on urban, or most of the emphasis was on the rural side. I mean, we said some of the implications for the urban model and how the data do in the urban setting.

So it's kind of a little hard to me to go back to that macro where we kind of don't have firms. Like it's unclear to me exactly what we're going when we're talking about individuals having projects. And, oh, they might employ some other people in the village. But that still sounds to me as like little things. And then how do we translate that into if we think more about it in a macro perspective where there's really no firms and [INAUDIBLE]?

ROBERT TOWNSEND: Well, the Kaboski paper it's true used the annual rural resurvey. So it's about village and village size. But, before we did that, we did the paper with Alex. And that's about obstacles. And there we used the annual data for both the rural and the urban or the annualized version of the rural. But, anyway, half of the estimation was about urban households.

AUDIENCE: Yeah, I remember that.

ROBERT TOWNSEND: And so--

AUDIENCE: And we said that the rural [? other ?] was very important. But it still kind of don't resonate with me like how-- like what are we thinking about when we think-- when we think about rural? Are we thinking about firms? Are we thinking about little resalers? [INAUDIBLE]

ROBERT TOWNSEND: There's a whole spectrum of size and scale of firms in these countries. And it's true that, in the rural areas, they tend to be small, small or medium enterprise, household enterprise. That's actually true as well in the urban areas. Most of the firms in these countries are small. Most of them don't even hire employees.

Now, when we're doing macro proper, we want to make sure that we also include the medium-large firms if we have the relevant data on them, but I'm not sure I see so much of a difference between urban and rural. For example, there are very active money markets in the urban areas. Credit constraints, in some respects, seem as severe or at least constrained, very high interest rates, money lending, despite the proximity of financial infrastructure.

So, yeah, you can study urban neighborhoods. I once did this in Chicago. I started looking at urban neighborhoods, a Mexican neighborhood, a Korean neighborhood, an African-American neighborhood, and just seeing the differences and whether the institutions and the data seemed consistent with certain outcomes. So, as far as the method goes, I don't see a big difference.

It's true that you're often persuaded by the data you have, and this monthly data is rather unique to the rural areas. So we've done more with that, but that's just a constraint. It's not some-- we're thinking about, god willing, we will gather these monthly financial accounts data in urban areas soon.

AUDIENCE: So, from my perspective, I think that, when we talk about rural and urban, if it's not Bangkok, urban isn't really like urban here. It's just like small towns. So I wouldn't think that like-- I would think that the results from the rural data [INAUDIBLE] generalize to urban areas. But when you think about-- I think your question is more like big, big firms or big firms in developing countries.

AUDIENCE: Yeah, so my point is that, sometimes, most of the time, like, when I think about TFP and studies about TFP, they have this concept about TFP growth like in a country, which is possibly unclear exactly what it needs. But it's like how much of what we're doing at the micro level then translates into this like super aggregate measure?

AUDIENCE: And you think that most of this may be coming from the large firm?

AUDIENCE: I don't know. I mean, I just-- again, like China is different, but like that one paper was about state-owned enterprises, which I'm assuming are big. Like that one made sense and was sort of-- which is why I kind of can't reconcile the two.

ROBERT TOWNSEND: The Jeong, Townsend paper is about TFP at the macro level in Thailand. And there we use socioeconomic survey data, which includes some data on larger firms. I suspect it's dramatically under sampling larger firms. And, nevertheless, we explained something like 72% of TFP movements at the aggregate level.

So I mean that was a real eye-opener that development in micro is all about-- largely about macro TFP. It's just efficiency. You know, TFP literally means nothing more than how much-- or, changing TFP, how much more output can you get period after period, even holding the level of capital and labor and so on fixed? So, to the extent that intermediation allows a better allocation of resources based on individual talent, you should expect it to show up in TFP.

AUDIENCE: You do have a new survey that focuses on big firms. Do you have any work in progress that maybe the result is different from what--

ROBERT TOWNSEND: Well, it turns out that big firms are harder to find in some sense. They're harder to get into. This TEPS survey in Thailand ends up looking a lot like the urban survey, even though it was designed to try to pick up the larger ones.

I mean, there is something called the missing middle, and I don't know whether it's missing because of the survey design techniques that are used or just missing because it doesn't exist. But the missing middle refers to the sort of life cycle of firms and the fact that we tend to see either the very big ones, as in surveys, census of manufacturers, or the very small ones, as in household survey data.

And the interesting thing is, where do the big ones come from. What is their history? Go back in time. Did people move out of villages into the town and set up their businesses there? And that's something we don't know as much about as we would like to know.

It's even hard, actually, to imagine how to sample to figure it out because everything is endogenous. You can comfort yourself when you have a fixed enumeration of households, as in a population census. Then you just sample at random. Or you have an official registry of large businesses, and then you sample them at random. So we can take comfort that we have an unbiased, stratified, random sample, but that doesn't get at this missing middle if-- yeah. Yes?

AUDIENCE: So I have another very general question about structural model. So we've seen a lot of models with this heterogeneous agent. So, obviously, there's all kinds of sources of heterogeneity. And, in your model, you only choose to model-- you cannot model all of them. And what you're going to model is what you think is playing a key ingredient in driving the mechanism of the model.

So, for the other heterogeneity, it's what? Like, in the paper with Kaboski, you just filter-- adjust your data by filtering the other heterogeneities. I'm just wondering whether like, for example, people think that there are three or four key heterogeneities that are driving the decision-making at the household level, but you're only modeling one of them, and you're filtering out the others will be getting a lot of [INAUDIBLE].

ROBERT TOWNSEND: Integrating up I would say, rather than filtering out. There is a question of the dimension of the unobserved heterogeneity that's in that model and in what we're going to talk about momentarily. You have-- like an econometrician, it's very similar in spirit.

You have the observed covariates x on the right-hand side, and then you have some unobserved household or other fixed effects on the right-hand side. And you can assume all kinds of things about those unobserved heterogeneities, but you have to assume something basically. It's hard to do something completely non-parametric. So you're making assumptions. I'm not denying that at all. And it matters what you assume, but it's not as though the heterogeneity isn't there.

AUDIENCE: So it's not about unobserved. So, for example, household [? that respond ?] in that paper, the heterogeneity is in permanent income and wealth, liquidity, wealth, and permanent income.

ROBERT TOWNSEND: Yeah, wealth might be observed. Permanent income is not observed.

AUDIENCE: And household composition is observed, but it's not modeled.

ROBERT Oh, now that's something we did for the data. And that's more a trade-off in methods. Let me draw an analogy. If you did real business cycle modeling, you would basically take a trend out of the data. I mean, in the US, there's just this huge trend. I mean, if you looked at the business cycles around the trend, you'd like what is this subject called business cycle. Of course, when you're in a recession, it seems real enough.

So people have a filter. You could take a linear trend out. You could use this Hodrick-Prescott filter and take out some kind of smooth moving average. And then, having taken it out, you'd just focus on the deviations around the trend.

So what we did was take out all the demographics, which is not pleasant either, but, I mean, it at least-- the other thing to do is to leave it in the data and use a model that doesn't have it. And that seems weird too. That's like doing business cycle modeling and leaving the trend in.

AUDIENCE: So, id you don't-- for example, but, if the question that you're-- like the model, like, if people think that household composition or demographics is playing some sort of key role in the household decision-making, but you're filtering it out--

ROBERT That's a fair criticism.

TOWNSEND:

AUDIENCE: So there are criticisms like that?

ROBERT Yeah, yeah, absolutely. Yeah, well, it sounded like all the research got done. Oh wow, we'll see you on Thursday.

TOWNSEND: No, no, those are good, productive thoughts, which is it's not like-- you can be sort of uncomfortable with research in general and the apparently arbitrarily way in which different researchers make assumptions.

But where research kicks in in a very positive way is to say, no, I don't believe that is a reasonable assumption. I think it would matter a lot. Let me try to incorporate it and maybe leave out some other things and see what difference it makes. That's where good research is born, absolutely. Yes?

AUDIENCE: So a lot of the models that we've covered has a wage as a key parameter. But, in like some research in, if I remember correctly, your paper in *Econometrica* doesn't find a effect on wage, at least short term. And, also, like Esther and Abhijit's paper didn't also find wage, like effect on wage, of microfinance [INAUDIBLE].

ROBERT Joe and I did find an effect on the wage in the astructural model.

TOWNSEND:

AUDIENCE: Oh, OK so that's totally [INAUDIBLE].

ROBERT Yeah, yeah.

TOWNSEND:

AUDIENCE: I missed that.

ROBERT But, again, that's a good example again of these trade-offs, which is you have a sort of a search algorithm through the lens of a somewhat reduced form, maybe instrumented variables to see what's big and what's small. And we found an effect on the wage.

Now it is true that in that model, the structural *Econometrica* paper with Joe, we did not try to make wages endogenous. In fact, we allowed investment, but we moved away from occupation choice and these key ingredients of a lot of these macro models. It's not like it couldn't be done. It was just something we chose not to focus on.

AUDIENCE: I guess my point is-- so I don't think there are a lot of evidence, especially [? treating ?] endogeneity of financial development. But it seems to me that, at least like RCT of like MFI and [INAUDIBLE], they usually didn't find, but I only know two papers, but they didn't find like effect on wage. So I'm kind of wondering like what's your [INAUDIBLE].

ROBERT TOWNSEND: I think, if you think about-- you have to think about the scale of the intervention, all right? So, if you have an RCT, even with a fairly big sample, in a small, open economy, it's arguably true that you wouldn't necessarily expect to see anything with the wage. But, if you were scaling that same experiment up to the national level, then you probably would see it.

Actually, that's a very good argument for doing both basically and, in particular, not getting misled by program impact. Ester's thesis in Indonesia was about education reform. And she was worried about whether the wages would go down, as you had more educated cohorts. So people have worried about this kind of thing for quite a long time.

There is some work in Canada. RCTs are a bit sort of like they look new, like people maybe haven't been doing them much, but it's not true actually. If you can go back into the '60s and '70s, there was work in Canada on I think it was either a pension or social security reform where they did exactly what I was saying. They kind of backed that impact locally, and then they extrapolated to what would happen to the rest of the country if it were implemented nationally with the structural model. Yes?

AUDIENCE: So this is a very broad question, but [INAUDIBLE] asked something about the financial [INAUDIBLE], especially corporate financial [INAUDIBLE] equity markets and money markets. So these markets of course are [INAUDIBLE] to the way economic surplus are divided into different people or different organizations. So, as a result of that [INAUDIBLE] impasse, how, for example, new projects or new policies have benefited different types of groups or organizations-- so there are some political economy issues related to the imperfection of the corporate financial [INAUDIBLE].

ROBERT TOWNSEND: Impact of the corporate finance on--

AUDIENCE: Structure on-- so political economy concerns about who benefits from new policies and new projects.

ROBERT TOWNSEND: Well, there's two things. On almost anything we've done, there is sort of the winners and losers. And that, in principle, that has nothing to do with political economy. That's just mapping out who's going to get the benefits. Now the next step is, if you think there are barriers, it may be that the losers are blocking reforms. And that's definitely a political economy.

AUDIENCE: Oh yes, then my question is does the literature seriously investigate the second step you just described, [? fair ?] wages.

ROBERT I mean, Daron has occupation choice models where, basically, the oligopolists are the ones sort of dictating the outcome. And they are a force for financial repression because they want to keep the wages low. I mean, that's at least a step.

We, in this class-- let me try to think if we actually covered it. A bit, yeah, you saw one version of that. We can model how financial service providers ought to be behaving if we believe the structure of the model. Then we look at the data, and we discover that there are gaps in some places and too many banks in other places.

And so, through the lens of the model, you sort of get suspicious that-- develop a null hypothesis that there may be political considerations at work. Now how explicit or evil or well intended it is, that's-- you have to go in there with your eyes open, but not necessarily assume.

AUDIENCE: I have another question if no one else does.

ROBERT OK.

TOWNSEND:

AUDIENCE: So--

ROBERT Anything but the lecture, huh?

TOWNSEND:

AUDIENCE: No, I just-- this is-- so I'm kind of embarrassed to ask this question because it seems like it's something that I should have been able to figure out on my own. But you see a lot of people throwing around the word structural in ways that are confusing and in ways that sometimes-- sort of I guess the natural and then from what we've looked at and all your papers, I guess I'd sort of be inclined to say your models are only structural if you are identifying like underlying parameters that are in a model of interest using econometrics or using calibrated simulation.

But I see a lot of people calling stuff structural that doesn't necessarily actually identify those parameters. Have I got the wrong end of the stick? Or is there like a nuance that I don't see?

ROBERT I don't-- I don't actually know for sure, depending on who's using the words, what they have in mind.

TOWNSEND:

AUDIENCE: Yeah, me either.

ROBERT I think, legitimately, many people are worried that assumptions are being made in structural models, which are never tested, and that you're seeing the observables through these auxiliary assumptions. And, if they were sort of totally off and totally wrong, then you're not learning anything. So that tends to push you toward data summaries. It tends to push you toward not assuming things.

But then there's a flip side, which is actually the lecture, which is this notion that there's theory-free tests. It's just not true. Even though RCTs and-- you know, they're all assuming some things that you might think are rather innocuous, but, when you start to think about the structure of decision-making, there's content there. And those assumptions could be wrong.

So I view the tension that way, which is trying to assume as little as possible, if you can get away with it, in terms of the underlying distributions of unobservables and so on, but, at the same time, mindful that data work without theory has big limitations.

And that's a debate, philosophical and otherwise, that goes way back, I mean, back to business cycles. You know, Burns and Mitchells-- Burns and Mitchell have this business cycle measurement, the NBER dates actually. And Koopmans came along, and he said it's just crap. And there was a lot of fighting.

So how can you measure something if you don't have a model in mind, et cetera? So I'm not saying there aren't good answers to that, but this is not just a debate in development economics. This is, again, kind of a judgment.

AUDIENCE: So it's a kind of vague question, but we've seen in the lecture a lot of models and evidence that financial [INAUDIBLE], [? let's say, ?] sort of like what financial development does to the real economy. And how would you compare that to the situation in developed countries? What is like-- what is the difference?

Because like a lot of models-- at least model-wise, you can apply it to developed countries. And, at first glance, like, for example, like for consumption, people aren't risk sharing quite well. So like it's quite close to this situation where the market is complete.

On the other hand, if you look at macro level, as we saw in the first lecture, like there's-- or like people tend to think that financial development [INAUDIBLE] varies over countries and affect [INAUDIBLE].

ROBERT TOWNSEND: Well, we've used the US as-- some of those papers used the US as like the benchmark standard, as if it had perfect and complete financial markets. But I think we all know now just how badly the US and other advanced countries have done over the last-- since '08, '09, and so on.

So, actually, what's going on in macro is this huge rebirth of interest in modeling the financial sector. There's no consensus on how to do it. And I think that literature would benefit from the advances that have been made in macro development. But there's no reason why-- I mean, the same issues will come up when thinking about the US or Spain.

I mean, Spain is an example, for a good example, depending on whether you think it's an advanced industrialized country or not. Spain liberalized its financial system in the late '80s with a series of reforms on interest rates and the nature of business, of savings and loans, and all these providers.

So these little regional banks, like Santander up there in the north and, basically, along with many of the others, expanded their financial infrastructure, establishing branches, spreading throughout the rest of Spain and, eventually, of course, on to Latin America.

And then now we have this mess, right? So what went wrong? Or what went wrong in the US mortgage crisis and so on? So I think those are-- there's hardly a consensus, as I said, but those are really important topics.

This is haunting countries, as well as the IMF and the World Bank and so on, now because countries want to grow. There's a lot of poverty, a lot of inequality. And there is this notion that expanding financial systems, finance, can cause growth, and growth can trickle down and help reduce poverty and so on.

So there is this sort of development view that finance ought to expand in some way and cause growth. And, at the same time, there's this worry that expanded, deeper financial systems are going to be prone to instability. And that's a force for repression.

And that was the first lecture of the class, actually, was how do we deal with both these, get both these things on the same page. But, for sure, today in the world, throughout Africa and Asia and so on, people are sort of wanting to expand and, at the same time, quite scared, as regulators, that they're going to take the blame in the future when something goes wrong.

All right, so I'm going to give you a 30-minute summary. Fortunately, some of these themes have come up. And, as always, the lecture notes are on the Stellar website. So it's OK if we don't cover everything in enormous detail.

Oh, by the way, one last thing, which is I did write out, for those of you who are worried about the generals, sort of a reader's guide to the-- not a reader's guide, a summary of the questions that we've been addressing in the class. It's just kind of a way to organize the material, and I'll post that on Stellar and-- who's also going to have some sessions. I think there's only four or five of you that are actually going to take the generals, but, if you are, it's easy to get lost in all this material. So I wrote up this sort of summary.

AUDIENCE: [INAUDIBLE] [? going to ?] need it for next year.

ROBERT TOWNSEND: Yeah, well, by then, research will have advanced, and we'll-- OK, so this is about this structural versus reduced form stuff. And the topic is the impact of expanding financial systems. So it's both a methods and a topic that basically is the hallmark of the class.

There's three things. I'll probably only barely get through the first. One is sort of IV versus structural analysis, all in the same model, basically. The second is this expansion of banks in Spain, which I was already mentioning, using what the IO guys do, the way they avoid having to spell out the entire structure of the equilibrium by using reduced form Markov processes. And then, finally, if I were to have time, I could tell you about the limitations of that. The point is there's no one thing, but you need to be aware of the strengths and limitations.

So this is a paper I wrote with Urzua, and it's about occupation choice and about unobserved heterogeneity. So a lot of this is already familiar. You've got sort of your wealth if you're a wage earner, your wealth if you're an entrepreneur. Instead of having just a flat wage for everyone, there's sort of this productivity shifter, which varies with households.

Or you can run a firm. There's some setup costs unobserved, again, varying with households, [? i. ?] The profits come from maximizing revenue less the cost of inputs. In this case, there's no financial sector. Ironically, I'll get there. And you just have to finance the setup costs and your capital with your initial wealth.

So you saw this. This was day two. This was the Lloyd-Ellis and Bernhardt model, versions of which we took into or modified in some of the other papers in the literature. So you get to choose one or the other. So we can create this decision rule, a dummy D for do profits dominate wages. If so, do it, and, otherwise, don't. OK, so it's an indicator.

Note that it depends on observables, the economy-wide wage and the wealth, which varies across households, and these unobservables, which are these setup costs or talent for both being a wage earner and being an entrepreneur. We're not going to see those.

So, as you know, it's sort of naive-- you know, incredibly simple, but, in some respects, powerful way to write down the observable depends on which thing you do. But, if D equals 1, then $1 - D$ is 0, and you have the profits and earnings of being a firm versus, if D is 0, you're going to have the wage earnings from being a wage labor.

So you can write down, if you were running a regression of earnings, you could let that regression have wealth in there and wages. And then you'd have this sort of dummy, which is a 0, 1 binary variable, plus some epsilon term, as regressions tend to have.

However, following the structure of this model, we know exactly what the epsilon is. It's this D -weighted average of the talent. Now you don't change your talent when you make an-- you have your talent in both of those occupations. And that's what's driving the choice, OK?

So, basically, the worry, correctly, is that D on the right-hand side is correlated with epsilon because that's self-selection, OK? You choose the occupation with which you're most productive, and part of that choice is based on unobservables. So we see it both in the choice and in the error.

OK, so we need an instrument. Let's imagine in this economy there's some subsidy or tax. And it's lump sum and administered, say, randomly. Well, then, clearly, the decision variable, which has a ϕ in it now is influenced by that. And, the higher is the subsidy, the more likely you are to be running a firm, other things equal, right?

But, because it's lump sum, this instrument, this subsidy, affects the choice of what occupation you're in, but it does not influence the outcome conditioned on choosing. If you choose to be a firm, you're going to get that subsidy or not. And that leaves all the other decisions about labor to hire, et cetera, et cetera. And none of those decisions, none of those outcomes, depend on the subsidy because that's sort of by-gones are by-gones. You can't do much about it.

All right, so we could imagine running that regression, but we regress the decision to be a firm in the data onto observables, like the wage and wealth, but also onto this subsidy as an instrument. That's called the IV, the instrumented version of the decision. And then we run the outcome function with the instrumented version, OK? So, hopefully, this is a review.

And what is the policy outcome? The coefficient on d after instrumenting it is supposed to be the gain to being a firm, as opposed to being "treated," quote, and being a firm rather than being a wage earner.

But it isn't exactly that. Or, at least, put on your thinking cap. It's the local average treatment effect. It is the average gain in income for those people induced by the subsidy to switch from being wage earners to being entrepreneurs.

So you're not getting-- you're not getting other things, which are easily, similarly sounding, like treatment on the treated. Treatment on the treated sounds like, oh, well, people became firms. What was their profits relative to the earnings they would have had if they were wage earners?

That's what this says, right? What is the average in the population of profits with the θ 's in there less wages, given that they made the decision to be a firm, and their wealth is basically b ? So I'm doing this wealth category by wealth category.

And this is the average treatment effect. This is what the average would be if the whole population became firms, taking wages as given and so on. So, basically, be careful because LATE is not necessarily equal to-- what you get with the IV is not necessarily equal to treatment on the treated and not necessarily equal to the average treatment effect. Is this a review for you guys? OK, good.

So you can put more structure as a way to back out the average treatment effect and so on. Now this is like assuming that those error terms are multivariate normal. And then you run this logistics regression, basically, binary yes, no, be firm or not.

And you can back out these variance terms, assuming that there's no-- that the talent things have say zero mean, or you could adjust for that. And then you can get expressions for-- so then you want to basically run your regression of profits onto those people who have chosen to be firms, controlling for their wealth and their subsidy. And, if you didn't, if you weren't careful and did OLS, you'd have a bias, but you basically use that first stage probit to get a propensity.

The structural model tells you what the exact formula is for the selection based on the unobservables. And you put that into the regression. You literally put this into the regression and run it. And it, quote, "solves for the selection bias." Yes?

AUDIENCE: How much more computationally difficult is it to use some kind of data distribution or like some other kind of-- because normal is a bit of a stretch as an assumption for a talent distribution.

ROBERT Oh, I think that's quite doable. Look, this is just meant to be an example, a classic example, but--

TOWNSEND:

AUDIENCE: Cool, so it's not too-- OK, sounds good.

ROBERT But, I mean, this question that you guys were asking, so here you've made-- you're able to get average treatment

TOWNSEND: effect and treatment on the treated, but only if you make some assumptions. And some people don't like that. They'll say, oh, you know, it's too structural. But, in order to get answers, sometimes, you need--

AUDIENCE: You have to make assumptions.

ROBERT OK, now, I mean, the most powerful way to say this is we actually see the counterfactual through the lens of the

TOWNSEND: model. We can see what the wages would have been for people who chose to be firms. By definition, they're not in the data.

And this is a generalized version. And I'll just say that this is like local average treatment effect, but really local, like just taking derivatives of how that difference between profits and wages is going to move, holding wealth constant, as you vary something like a propensity score. And the propensity score has, not too surprisingly, everything to do with those unobservable talents.

So this is hard to implement. This one is really not easy. Josh doesn't like it, among other things, but he's not disagreeing that it's wrong. He's just thinking it's just not as operational. Now the nice thing, if you could back this out, then you can ignore this page, but, basically, treatment on the treatment and the average treatment effect are just sort of integrals of these marginal treatment effects. Yes?

AUDIENCE: So Josh sometimes recommends to apply LATE to non-compliance based on observable characteristics. So, in the typical data set in the context of occupational choice, what can we observe among characteristics of [INAUDIBLE]?

ROBERT TOWNSEND: So here the model is taking a stand that we're going to see their initial wealth. We're going to see the factor prices that they faced. We're going to see the profits that they make. And we're not going to see these unobserved things that are driving fixed costs around.

This is not the only model around. We're going to use the model to make these examples. I haven't gotten sort of to the punchline. I'm happy if it's a review, but-- OK, so we have the model, right?

So, if you don't believe this miracle, let's simulate the data from the model. Instead of looking at some of the cross-section that got treated with the subsidy and others that didn't, let's take a particular individual in the model at given parameters and see what the behavior is, in terms of switching occupation and so on, as you turn the subsidy on and then average up in that wealth category. So believing in the miracle is basically the math that you're going to see the average of this model-generated simulation of impact will be the IV. It's going to be LATE under these assumptions.

So here is the IV estimate. And, roughly averaging over wealth, et cetera, it's 0.45. OLS is-- by the way, OLS is terrible, but I guess you don't need to know why. There's nothing productive about the subsidy. The subsidy is just inducing people who should be wage earners to be firms.

Of course, the impact is negative. And OLS estimates it to be positive. Why? Because, in the observables, it's the talented people who are firms, but they're not being induced by the subsidy. They're already firms, right? I couldn't help myself.

OK, and this is the model-generated treatment where we actually have the panel. And you can see, numerically, it's quite close. All right, so that's reassuring.

Now let's go back to the topic, which is what is the impact of expanding financial infrastructure. Let's invent a village fund. Let's let banks establish branches-- forgive me for the moment-- at random. Like maybe the government told them something, or I'll get to that, but, for now, random.

So, if we observe these travel times or the effectiveness of local financial infrastructure, we can call this cost here to be an instrument, right? So, yeah, so the idea is just like the subsidy. The lower is this number, the more inclined you are to incur the cost to be able to do the banking.

OK, we rewrite the problem. This is the part of the model with financial infrastructure. They get to borrow money to cover the capitalization and the fixed cost. Wage earners, by the way, also get to put their wealth in the bank and earn money, interest rate on their savings. You've seen this model. This was, again, the second lecture.

And now we have the observed outcome. If you are in-- if you chose to be in the intermediated sector, you have another occupation choice problem within that sector. You can be a firm or a wage earner. These D's denote that choice.

Notationally, it's got a w and r in there. That r is a clue that we're in the intermediated sector because there's an interest rate. Because, if you were in the autarky sector, there's no borrowing and lending of funds.

And, of course, this is the occupation choice we had before without intermediation. So we're going to get sort of the observed earnings in the autarky sector and observed earnings in the intermediated sector. And, within both sector, there's this other margin going on, which is the occupation choice.

All right, now we're reaching the point here. We have a double margin. There are two things people are choosing from, whether or not to go to the bank and what occupation to have. Sounds like it shouldn't cause a problem.

This is basically-- I'm sorry this is so hard to see, and it must be even harder for you than me. I'm right next to it. This sort of symmetric Y is like an indicator variable for choosing to be intermediated. These Y's with a little bit of tail off to the right are incomes under intermediation and autarky. It's this way in the paper, and somehow the slides got created this way.

But, anyway, so the gain is-- let's skip that. So what can I say? It's a double margin. So you get to choose whether or not to go to the bank. And, within that, each category, you get to choose your occupation.

We can linearize the profits functions to be functions of the observables and the unobservables, substitute in those decisions, and write out the expression for what you see in the data, namely, this income, but it's just a weighted average of all the choices, depending on whether these D's are 0 and 1's or these symmetric Y's are 0 or 1's. That was a bit of a rush, but, I mean, I think, conceptually, you understand where this is coming from.

But, not only that, look at this error term. Oh my god, the error term is complicated. And it also has all these decisions, but note it's got the occupation decisions in both sectors. Depending on what you do, certain terms are going to kick in and kick out.

There's massive selection going on about whether or not to go to the bank, based on what occupation you're going to choose, conditioned on going to the bank or not going to the bank. And we don't see any of this. We just see the choices. We don't see these error terms.

So what can an instrumented version of intermediation give you? It can give you the local average treatment, the income gains that are a consequence of a lowered cost that allows you-- makes you want to go to the bank, as opposed to what you would do if those costs had been higher, and you're not going to the bank. So the local average treatment is the impact on earnings that have to do with those people who are induced to join the banking system as a consequence of the placement of branches or varied costs of access.

Now you have to be careful in terms of what this isn't. For example, do we want to look-- what do we mean when we say we want to measure the impact of improved financial access on the profits of entrepreneurs? You want to do an IV for that? Good luck.

Now what can go wrong? Remember monotonicity and independence, OK? So monotonicity means that more people are inclined to do something as a consequence of the policy variation.

Are more people inclined to be firms as a function of having reduced cost of financial access? Maybe if they were credit constrained and then make a lot of money in business, but there's another subset of the population, those not so talented people who are in financial autarky, who would be actually induced to leave business and put their money in a bank. So it's not monotonic. The effect of improved financial access is not monotonic on things like the income of entrepreneurs.

You can, if you're careful, think about what is the impact on profits of entrepreneurs, of those people whose decision to be an entrepreneur is not affected by financial intermediation. Now that's an interesting question, but it is not necessarily the one that we had in mind to begin with.

And who are those people? Well, it kind of depends on their talent, right? So you're making statements about where these talent cutoffs are when you're conditioning on, quote, "things like people who were not making their"-- if you just run an IV in this cross-section, just be mindful of what you want to do with the occupation categories.

Now this is where the rubber hits the road. Am I just talking fancy theoretical models here? I don't think so. This is just the most innocuous kind of unobserved heterogeneity. And yet it tells you that IV is not a panacea. You have to be very careful when you use it to think about the questions you're able to answer.

So let me just jump. So, again, we have the model fortunately. So we can generate any data that we want from the model and talk about occupation choice or talk about financial intermediation.

So, at the parameters we put in, which have to do with that LEB thing you saw in the second lecture, roughly, we can talk about moving people in the model from autarky to financial intermediation, in other words, changing their choice problem, damning them to autarky, or allowing them to be in the intermediated sector-- that's the conceptual experiment-- and looking at what happens to people by occupation.

So we've got guys who are wage earners before and after. They're not moving. And we kind of get their gains. We've got people who move from the wage-earning category under autarky to entrepreneurs under financial intermediation, and you'd like to think about those as the talented poor people who now are able to borrow.

And then there's this negative number. And I looked at this this morning, and I'm like, [? oh, ?] what's wrong, negative. And then I remembered, oh yeah, the subsidy. So the subsidy is in place here. So you can actually get adverse selection, if you want to call it that, of people who really shouldn't be moving from wage earner to the entrepreneurial category.

And it's precisely those people that get impacted on the margin. These are local average treatment effects. These are people who are induced to change their behavior as a consequence of the program.

Then we've got people who are firms no matter what, and they gain. And, people who are from entrepreneur under autarky to wage earner under intermediation, so these are the not so talented guys who basically do switch and do gain as a result of the program.

So, I mean, so you might think that, at plausible parameter values, you wouldn't see any of these occupation switchers, but you do. And, again, the models that we've been talking about sort of assume something like that's going on. That's where the TFP is coming from, except we didn't have a subsidy. Well, maybe China did. Maybe we were removing the subsidy.

OK, this will be even shorter, but it's a nice review in a way. You remember that Greenwood and Jovanovic, the forward-looking model? People solve stochastic, dynamic decision problems to maximize expected utility, choosing how much to save and where to put their money in terms of safe or risky activities or assets.

We're going to put a little heterogeneity in this discount rate. It could be other places. Now we're just going to experiment and see. So here let β_i be equal to some common average, $\bar{\beta}$, and then some heterogeneity, unobserved heterogeneity, that makes β_i different from $\bar{\beta}$.

And then we saw this optimization problem. We have savings, fractions of wealth to put in sort of safe or risky. Or, in this case, I generalized it to entrepreneurial and wage-earning activities. You have aggregate shocks in both. You have idiosyncratic shocks in both. You have a law of motion for wealth. Wealth today, fraction saved, fraction put into risky and safe gets you wealth tomorrow.

We have two sectors here, intermediated and non-intermediated. The non-intermediated sector is like this financial autarky sector. So we're going to look for this value function, W_0 . If you're in the intermediated sector-- oh, sorry, well, I'll do the next one. If you're in the intermediated sector, you're looking for another value function that solves dynamic optimization.

Now, for each one of these, if you were there and destined to be there forever with these constant relative risk-averse utility functions, you have some simple sort of facts like the savings rates are constant. Savings rates depend on β , but give me the β , and I'll give you the savings rate. And that's true in autarky, and it's true as well in the financial intermediated sector, OK?

But, because it affects the savings rate, it's going to affect income. And so part of the unobserved error is going to be the product of income with this θ_i , but θ_i was that preference parameter we don't see. You with me? Hopefully, you can at least conceptually see where we're going.

OK, and so there's-- similarly, there's a savings rate and consumption for the intermediated guys. And then you have the big choice here. This is Greenwood and Jovanovic. So there's this cost Z of joining the financial system. And, if this cost Z is sort of a random variable in the population, it's going to affect the choice. The issue is whether V is greater than W .

So, if all of this choosing were happening only in the initial period at t equals 0, then I've got a valid instrument. I have something that affects the choice of intermediation, even in this dynamic model. And then you're in one sector or the other.

The savings rates depend on the error term, which depends on θ , but this occupation choice here-- sorry, this this participation choice is instrumented with Z . And Z has nothing to do with the unobserved preference heterogeneity. So I have a valid instrument. I have monotonicity, OK?

However, that was not the way we dealt with that model in the lecture. We talked about endogenous financial deepening. That makes households forward-looking, planning, potentially, to join the financial system at some future date. Well, now we're in trouble, even with the same instrument, because what we see prior to the selection is the endogenous evolution of wealth. And these guys will be saving up to cover those fixed costs. It's no longer true that the savings rates are just constant fractions of income.

What does it mean for RCTs? It means you should try to keep the control population in the dark about the program because, if they know it's coming, and they'll eventually get treated, they're going to change their behavior if they behave the way they do in the model.

So Ben surprises people with the road construction when he dug up the dirt to see what the pavement was made of. That happened after the fact. He didn't tell them he was going to do beforehand for obvious reasons.

But it's very hard in the field to maintain this because people want to know why they're answering all the questions in the questionnaire and not getting the program. They're not even supposed to know about the program. You know, there are ways to try to fix it.

Now it's true that, if you had like a surprise, even though they're following this dynamic model, you say, oh, today is we got a special deal. It's not going to be repeated tomorrow. Then you have this one-time-off experiment. You have this Z that's lowered that they got surprised, that prior behavior isn't influenced by it, that will induce selection and so on.

Well, what am I doing? I'm just reviewing, in a way, what we've already done with these kinds of models. This one is dynamic, but putting it in the language of when RCTs and instruments will work and what to be kind of leery about. I honestly don't see how you can talk about IV or RCTs, random as they may be, without having some framework in your head to think about the model.

And the rest I'll skip. It's about that financial expansion in Spain. I'll tell you the key insight the IO guys had, which is, basically, you worry about an equilibrium where everyone is anticipating everybody else. You say, well, hey, if you believe it's an equilibrium, then it ought to be in the data. So you start running reduced form regressions of how your competitors behave as a function of what you did. And that's kind of the forecasting aspect, which you can bring into the individual problems.

I know it sounds a bit counterintuitive, but, basically, the reason it works is because everyone is getting these shocks. So no node-- you know, you can be at any node and go to any other node because there's always a shock that's going to rationalize that choice. It's like Hutz and Sedlacek way back when. And it's the foundation of these discrete choice models.

Now that doesn't always work. And the last thing I won't even tell you is a situation where it might fail. So the point was not to drive nails in the coffin of one way of doing business. The point was to have an open conversation and have everything on the table and so to increase our sensitivity.

If you don't think heterogeneity is-- unobserved heterogeneity is a big deal, that would help you a lot, but, when we looked at the impact of microcredit, we saw a lot of heterogeneity in outcomes, whether it was Hyderabad or Thailand or Morocco. That seems to suggest this selection on unobservables is quite important. OK, good.