ROBERT TOWNSEND: We want to talk about theoretical models. They tell us stories. The key variables are very much of interest-- growth, inequality, poverty, but just telling a story as elegant, as it may be, even with surprises does not even necessarily tell us that it matches up to reality, and therefore, one might be a little bit cautious in terms of formulation of policy.

Now it is true that having sort of a good conceptual foundation, a good well-articulated theoretical model is a very important step. I'm not demeaning that step at all. What I'm trying to say is it's a very useful first step, but in the literature we're going to review today, we've gone beyond that, and to try to take the next step and see how well it fits.

And how are we going to, quote, fit it? We're going to take the assumed micro-financial underpinnings and estimate those underpinnings, and get the bulk of the key parameters of the model pinned down in that way from the micro data. There'll be a few remaining parameters that we can't get from household choice or business choice, and we'll calibrate those.

And then we'll simulate the models and get what the model would predict for growth inequality-- income, savings, labor's share of GDP, and actually also some of the key underpinnings of each model, and the first one it's going to be the fraction of households running businesses and the second one is going to be the fraction of households participating in the financial sector.

So what do we do with these? Each one, then, after we get done looking at the goodness of fit, we can do welfare analysis of policy intervention. The first model has, to make it easier, an exogenously expanding financial sector at exactly the rate we see it in the data, and that's going to create winners or losers.

As the financial system expands, we'll get changes on the extensive margin because some households will go into business. We'll get changes as wealth accumulates in the amount to take and invest in business, and also because it's going to be a general equilibrium model, we're going to get changes in equilibrium wages and interest rates which have effects even beyond the direct effects, often quite big.

In fact, there's a debate, depending on which political position you take, about trickle-down, which is sometimes taken negatively, trickle-down meaning only the rich benefit from various things and the poor-- workers are lucky to get anything, a better way to say it is a rising tide raises all boats. The model kind of takes a stand on these things, and in particular, the wage, when you have financial deepening, can go up, and that is an enormous benefit to poor people, even unskilled poor people who have better employment and better wage opportunities.

The second welfare experiment is when, in the second model, the government takes over the banking system. I alluded to this at the end of last time, and again, when they liberalized this system and turned it back over to the commercial sector, that also creates welfare gains, which can be quite large, and we can-- the point is, we can use each of these models, if we believe the model and we believe more or less the numbers we put in for the parameters, we can quantify these welfare gains and losses.

There's some corollaries that come out of this. First of all, growth doesn't necessarily have to go up as the financial system expands. We reviewed some literature last time about finance causes growth that seemed plausible and reasonable and at least at a reduced form since it seemed to fit the data.

But in the second model, as opportunities expand, you can create endogenously more financial infrastructure, but it comes at a cost. And just like anything else, investment not may not pay off right away. So some resources, quote-unquote, disappear in this creation of financial infrastructure, and so in effect, the growth rate levels go down and growth rates are slower for a while than they would have been without the liberalization.

Another caveat coming from the model is, both models today are transition models-- that is to say, they are not yet in steady-state, although if I remember, I'll try to say something about the steady states of each of these. And so in some respects, the models are not generating stationary time series. Implicitly when you run a regression, OLS and so on, you're kind of assuming you're getting data drawn from as stationary time series.

So we run regressions on some of these model-generated data where we know exactly what's going on, and sure enough, you get kind of spurious, bogus, or erratic coefficients on the very same regressions that the literature has been running with actual cross-country data. So it's sort of a caution.

Now toward the end, there's a tendency to, in any class, any lecture for any course, you see models and you'll see, oh, there's a model, let's learn how to manipulate that one and maybe do a few extensions of it, or a second class with a second model. Toward the end, I will deliberately try to engage you in a comparison of these two models to make the point that the real research that goes on is trying to decide not just how to an extent a given model, but sort of what we learn overall.

It's not typically published, this stuff, actually. Mostly these conversations take place in offices and hallways or over beer and so on, but it's actually a big part of the research process. So as an introduction overall, a few words from a paper I wrote with Hyeok Jeong, which repeats all what I've said already. Basically we're not grabbing parameters out of thin air, we're going to estimate them with the micro data. We compare the predictions of the model with the actual dynamics.

There is a well-established tradition in other sciences that has to do with model validation. So it's quite common to basically calibrate the machine or the model or the process, and then sort of simulate and see how well it works without being recalibrated. Or put another way, if you have enough degrees of freedom and enough parameters and you're constantly tinkering, there's a tendency to be able to fit stuff.

You've seen this probably in your first course in statistics and ordinary least squares. If you run out of degrees of-- you can fit-- align two points, you can perfectly well-- or you can fit a polynomial even with nonlinear data, but then out of sample, you want to predict some more and it's just crazy, it's just completely off.

So here, we followed the tradition of kind of not-- deliberately not using all of the data. It is a matter of choice, though, what data to use and what to set aside, and I could imagine more, quote, experiments than we actually did. So there isn't a fixed rule for how to do this, and it's not even as commonly used in economics as I think it probably ought to be.

And I guess you know the main theme today is this marriage of theory and data. Frisch very early on in the '30s was a big champion of the mutual penetration of quantitative economic theory with statistical observations. I think that's actually on the econometric cover, if not their website. And in particular, we're somewhere between theory without measurement and measurement without theory. A big controversy in business cycle analysis along that line, a big debate between [INAUDIBLE] and Burns and Mitchell.

OK. So we're going to do these two papers that came from the literature. It's kind of a plus and a minus. One is Lloyd-Ellis and Dan Bernhardt, and the other is Greenwood and Jovanovic. Each tell a plausible story about what lies behind the growth process and the relationship between growth and equality, financial deepening, and those other variables that I listed. And we're going to try to estimate those models.

Well, the negative part of this thing is we learned how to do statistics and econometrics in part through the lens of these models by realizing how problematic certain parts of those models are. And I would not start this way if I were starting over. And as I go through the lecture, I will try to point out which particularly worrisome or problematic parts I have in mind in terms of taking these models to the data. But anyway, again, back on the positive side, that's literally what we did. We adopted wholesale, without any modification or qualification, exactly the equations that these theoretical authors used in their models.

Uh oh. That's not good. So Lloyd-Ellis and Bernhardt, LEB for short, there are three occupations-- farmers, workers, and entrepreneurs, and it's costly to set up a firm. But these costs are like talent, they vary with the individual. The more talented you are, the less costly it is.

It's assumed that the cumulative distribution of talent is distributed according to capital H here where x is the level of cost. And you have to be careful with the notation where we're talking about densities versus cumulative distributions, so just kind of bear that in mind.

There's also an initial distribution of wealth that starts off the model, G of b where b is sort of bequest. Just to alert you, this model has a relatively naive dynamic part. People essentially live for two periods-- they eat today and they leave money for their kids or their tomorrow's next period as if leaving a bequest, hence the notation b for a Bequest.

So what earnings two people have in a given period? Farmers earn kind of a subsistence living in farming, and they carry their wealth over and have it at the end of the period. Workers earn wages-- excuse me-- and entrepreneurs make profits. Everybody has this initial wealth which can vary across individuals carried over to the end of the period, essentially. A little bit of an exception for the entrepreneurs.

When Dan wrote his model, he kind of had in mind an earlier development literature which makes agriculture out there in the rural areas versus factories and manufacturing in the cities. So they had in mind a kind of cost of living adjustment, so basically workers earn their wage, but they have to pay this cost of living, which gives a little bit advantage to being in the subsistence sector.

Actually, you can see it right away down here, if sort of in equilibrium to have both farmers and workers coexisting, these two equations have to be equal to each other. So as long as the wage does not exceed-- the wage less this cost of living does not exceed gamma, then a household would be indifferent about whether to migrate to the city for a wage and pay that cost versus stay in farming. Yes?

AUDIENCE: So is this going to be abstracting all the way from land markets?

ROBERT TOWNSEND: Yes it is. And-- OK, let's get to the firm. Well, let me just follow up on that. The model as estimated and calibrated will actually deliver the fact that for fairly long periods of time, the wage stays relatively low and you have all three sectors coexisting, although there will come this knife-edge time where the wage starts to pop up, and then quite unrealistically there should be no more farmers.

You would think this model is doomed to fail with that kind of knife edge, but surprisingly in Thailand, there are still farmers, but the wage did stay relatively flat for quite a long time and then grew rather quickly.

What about firms? So firms maximize basically profit. The price of the output or the consumption goods is normalized to 1. So firms have output, which is produced using capital and labor. Labor can be hired. Those are the workers, you have to pay them a wage. And capital, basically, is coming out of initial wealth. So capital and consumption are kind of 1 to 1.

So you have your sort of initial wealth, and the capital has to be financed by that. Capital cannot exceed that amount. And actually, there's this fixed cost, and that also has to get paid. There's a lot going on right here, because there's two constraints. One is you have to pay x to get started, and the second is that you can't borrow at all. All the investment is basically self-financed.

Now that said, you could imagine kind of weakening this a bit and allowing the amount you could be borrowing to be some multiple of a b, and next class, we'll talk about many models that do something exactly like that. But for now, no credit at all. So this just-- these costs just subtract right off of output as if at the end of the period, or b minus these things is what you have at the end of the period. Then you get you get the output and you pay the workers.

So the key in this paper and so many other papers is an occupation choice map that really summarizes the possibilities. So the two key variables from a household's point of view is the initial wealth that they have at the beginning of the period-- it could be high or low-- and the setup cost x to running a business.

And naturally, if your wealth is high and your costs are low, you're a talented person, you're, say, not only likely to go into business, but that finance constraint is not going to be binding either. You're going to run the marginal product of capital down to 1, and you still have more money left over, so you kind of carry it off to the end of the period.

There is a region of lower wealth, but yet somewhat talented people who are constrained entrepreneurs, they would actually like to employ higher levels of capital k than they are able to finance out of their own wealth. They would borrow if they could, but there's no credit yet. And actually, this straight line is when they can't even cover their setup costs, or on the margin, they're just able to cover their setup cost x out of their wealth b, and there's nothing left over, so they can't even run the firm on the intensive margin. And then the rest of the guys, they're kind of workers or agriculturalists.

So the shape of this is intuitively reasonable, right? The higher is wealth and the lower is talent, the more likely you'd be a firm or an unconstrained for a more-- the lower is your wealth and the more likely you are to be a worker. And the thing is, even when your talent is low, if your wealth is quite low, you can't be in business. These are potentially profitable entrepreneurs who are constrained on the extensive margin at these parameter values.

Now you say, what parameter values? Well, I drew some lines. The point of the micro estimation is to actually be able to draw those lines based on the micro data. So-- oh my goodness. I don't know how often this is going to be. So let me just grab three equations and let you focus on this.

Here are the functions, the structural functions of the model. There is a trade-off between consumption and bequests under this utility function. It's basically top Douglas in today's consumption and tomorrow's wealth. This little parameter here is omega. Omega is basically going to determine the savings rate. And we're going to have to calibrate it.

This production function is ugly. It was meant to be general, which sounds good. So it has capital, it has labor, it has the square of each of them, and it has an interaction term. Unfortunate-- so it's like translog. Approximately-- it can approximate almost any function reasonably well, but the problem is, it carries with it five parameters. Alpha, beta, xi, rho, and an unreadable script, I'm sorry. I thought it was sigma, but it doesn't look like sigma.

And this is one of the regrets that I have about pushing so hard in these models, because it turns out in hindsight, we generated data from that production function knowing the model, and it's still hard to recover these parameters. That's something you should get into the habit of doing. It's hard enough with real data, and you already know, the actual world is not like your model, it's an abstraction, but at least you ought to make sure that-- and that you can't really do too, too much about, because you have to make decisions about your model.

But at least you should make sure your estimation routine works well if the model-- if the world were exactly like the model and generated the data that you're about to use. Anyway, we did the best we could on that. And this is this distribution of setup costs, the cumulative distribution. The easiest way to disentangle this is to say, suppose m is equal to 0, then this disappears, and the cumulative distribution is just rising linearly in x. In other words, the density is flat. So this is just a standard uniform distribution.

That's also kind of nice on the diagram, because although I don't want to enlarge this, basically you want to know how many people of a certain x there are given a certain b how many people of a certain x there are before you go from firms to workers. And that's just the length of that line, and so a uniform density is convenient for the sort of-- the geometry of that figure. However, if m is you know more than 0, then you get that squared term. It just means the distribution of talent is skewed to the-- really, the distribution of cost is skewed to the right, so there are some very untalented people, and conversely when m is below 0.

Now all this big to-do about micro underpinnings, what do we see in the data? Well let's imagine we see the wage, and we also see sort of the net worth or wealth of the household, and we know their occupation. So what does the model predict? The model predicts that the occupation is, say, firm-- or yi equals 1 in a binary sense-- when that distribution of cost is-- this is-- you should read this as an x-- is on or below that partition between workers and firms. So you add up, as I was just doing, all those households with talent less than that.

Likewise, if you're not going to be yi equal to 1, you're going to be yi equal to 0 and a wage earner, and that's the rest of the residual mass or 1 minus this probability. So what is the likelihood? So just quick review, imagine you have a coin, and usually coins are fair coins, so-- I'm not sure about the Super Bowl-- equal probability of being heads or tails, and you see a bunch of data, 4 heads and 6 tails, and you say, well what were the odds of that?

Actually, it's a bit more interesting if the coin is not fair, because then it's not probability 1/2 for everything that turns up. But you just basically take the data, which is heads, tails, tails, heads, and so on, or a worker, firm, firm, worker, and start adding up the probabilities. They're independent events.

So this says log here, but if you undid the log, you would have the probability of being a firm, and then this yi would just be the 0 or 1, and when it's equal to 1, you're just getting the number of, say, heads or the number of firms in the sample, and likewise for the number of workers.

It's very convenient to take the logs-- and it takes this form because logs bring the exponent down in front of the variable. It's a cute trick, too, because you want to maximize the likelihood. Logs and levels or monotone in each other, you find the max of one, you're going to find the max of the other. So econometricians usually maximize logs for reasons like this.

And it turns out with all that algebra, those particular functional forms, for any set of parameters in the production function-- alpha, beta, rho, sigma, xi, et cetera, right? You can draw all those lines, and therefore, determine all these probabilities. So for a given guess about the parameters, you'll have a certain value for the log likelihood, and then you just maximize that log likelihood by choosing those parameters.

Essentially the data come first there on the diagram, you have sort of wealth-- thousands of data points, but you have wealth. And you don't see talent, that's an unobservable, but you do see their occupation. So you're moving those curves around to try to fit as many data points as possible.

Here are the estimated parameters. Alpha, beta, xi, rho, sigma, et cetera, and it's also estimated naturally, because that's kind of telling you whether there's a large mass of high or low cost people in the population.

AUDIENCE: So the line of matter is just the boundary between [INAUDIBLE] right? Divide [INAUDIBLE] doesn't matter for--

ROBERT TOWNSEND: Not in this case, no we didn't. Now if I had from the data households answering-- or firms answering the question, are we constrained? And you could even get more sophisticated. Running businesses, would we like to invest more given this shadow price? I've asked questions like that. Alp in the macro lecture half an hour ago was describing something like that in the US for firms right before the-- during the financial crisis.

And you could even do something on the extensive margin in principle. You're worried that those are lousy questions, like are you constrained? But anyway. It seems to actually work better in practice than you think it might in reality-- in the classroom. Calibration, not everything can be gotten from those occupation choice diagrams. So-- and actually, we changed the model just slightly. We're going to allow a very slight growth rate in subsistence income as if there were some technological progress in agriculture.

That savings rate is not estimable from the micro data the way we were doing it, and neither is the subsistence income in farming. The level of subsistence income we get from a socioeconomic survey done in Thailand in 1976, basically matching to the reservation wage, there is a slight growth-- very small trend over time in that wage rate, so we let it grow exogenously at 5.5% per year-- less than 1%.

And the Cobb-Douglas says the savings rate-- and again, we're going to get-- we're going to set it high, basically, at basically 1/4, but that matches evidence-- and you'll see this in China and many of the countries you're about to study, the savings rates are very high in these countries. So that's actually realistic. Learning by doing.

So from the original Bernhardt and Lloyd-Ellis paper, you begin to see a little bit of the dynamics and also their take on the development literature at the time they wrote the paper. First of all, why that myopic savings rate? Why just care about today and then tomorrow? Well, it just helps enormously in computing the solutions to these models.

It's like every day is a new day. Here's your wealth. How much am I going to save? That's a number, that's the fraction of wealth you're going to save, and the rest of it you know you've got available for consumption. So after you make your occupation decision. So every period, you kind of have like a static model where you're deciding on the occupation. How well you do, et cetera, depending on your occupation choice and your talent, determines your end-of-period wealth, and that you carry over to tomorrow. And tomorrow's a new day, the whole thing starts over again.

No, I don't believe these guys only live two periods. Although you will see, this tension in the literature between realism and tractability, people still use quite heavily overlapping generations models, and some-- and I mean like with three periods. So it's kind of hard to match year-to-year data that way, but you can tell interesting stories about the distribution of wealth and the impact of credit constraints. So this is an extreme version of that.

This is like Rostow's take-off, you remember? He worked for Kennedy and-- so this-- he says at first nothing much is happening, and then you get in this transition, wages begin to rise, but income and wealth are still growing first order, and then you get the advanced economic development and wages take off, and incomes and wealth basically really, really become quite high.

And I'll forget to say it, but if you let this model run long enough, it will converge to a steady state in which there's no more any growth in wealth. It's no longer a growth economy. And that steady state, because of this weird assumption about two periods and these costs, you will see people making transitions. They will have a certain wealth and they're going to be talented or not talented in certain things, and they will decide to, say, forget what they were doing last period and do something new today.

So you'll see a lot of churning in the occupation distribution, but you will not see any trends in anything in their steady state. This is going to be a recurrent theme when we go through these papers as to which ones are essentially in steady state, and maybe we're looking across a cross-section of economies that vary in some way, each of which country is in a steady state versus transitions. Today is about transitions, and I'm not going to spend a lot of attention to this steady state of this model, which is probably relatively uninteresting.

OK so now we get to the financial deepening part. So what we do is create an exogenous financial sector under which everybody's intermediated. It's like one of these dual economies, except for us, it's like you have free banks and access or you have zero access. It's very extreme.

In the banking sector, so to speak, it's not just about credit, it's about lending. So there's an equilibrium interest rate that's going to clear the market. More savings is equal to borrowing. And now you look at the return to farmers, farmers have their subsistence income, but they put their money in the bank, so they get 1 plus little r or big R rate of return on that.

Everybody-- when you have complete-- there's no uncertainty here, essentially. So you have a neoclassical separation between households and firms, which essentially means the financing decisions of firms have nothing to do with households. Another way to put this is, just put all your money in the bank and then borrow what you need if you need to borrow it all.

So all these guys are putting their money in the bank, including the firms, and then they borrow back what they need. Now it is true that the borrowing and lending rate are the same here, and that's how on getting away with this trick. Two lectures from now we'll have a paper focusing quite a bit on the spread between borrowing and lending as a measure of the cost of intermediation. This paper doesn't have that, this paper just assumes perfectly costless intermediation, and these rates are the same.

OK. So then alpha fraction of households are in this intermediate sector at this date. So how much is in the savings accounts? So you take the total population, alpha fraction of the population intermediated, everybody's putting-- in that sector putting b in the bank. So alpha times b is the saving.

And what's basically the borrowing level? Well firms, if you are firm, you're going to borrow to finance both your setup costs and your capital. So basically we add up or integrate up over the two characteristics-- that is to say the setup costs themselves as well as wealth.

You will only integrate up over firms when x is below some threshold. Now this threshold no long-- I should actually have drawn it on the other diagram. It's a completely flat line. Whether or not you're a firm or a worker-- I said this-- has nothing to do with your wealth, it just has to do with your cost. So if your costs are high, above this threshold, you're not going to be a firm. And if it's below, you are. If you don't have a lot of wealth, you're going to have to borrow a lot. So this is sort of the demand for funds. And then basically you're going to set the demand for funds equal to the supply for funds and find an interest rate that does that.

There is an inequality here only because, again, the 1 plus little r could go to 1, basically, so the cost of funds is basically like storing grain in the backyard, in which case there's some indeterminacy you might as well just carry it over rather than put it in a bank. But interest rates can be higher than that.

And likewise, employment, we've got-- we just integrate up the demand for work. Employment is tricky. People have jobs. You're either going to be a farmer or a worker or an entrepreneur. There's only so many people in the population. So if the size of the population is normalized to 1, you've got basically the number of entrepreneurs and the number of work workers.

The number of workers, it just comes from the employment of the existing firms where l is the labor hired of a firm, facing R is the interest rate, and w is the wage. There's two sectors here, intermediated and non-intermediated. They're different from each other, but the labor market is perfect. Again, not so credible, but makes it much easier.

So you can work for any firm, there's not even any real migration cost-- maybe that cost of living costs, and that's it. So you'll be indifferent at a given wage where you work. So you would just add up all the employment, add up the number of firms, and again, lots of the time, this thing will be equal to 1, you used up all the people, and you'll have to find a wage that equates the demand and the supply for workers.

So on Friday, Yan will kind of talk you through a bit more. Easy to say, maybe not so easy to do. Even though it's a static model, you have to find these equilibrium interest rates and wages that clear the market. You need an algorithm for doing that. You need to have a good guess about how to search and how to iterate.

AUDIENCE: Question?

ROBERT TOWNSEND: Yes?

AUDIENCE: This may be in all of these. So I want to think about investments in human capital. Should we think of that as-- that goes into x or should we think of that as a different kind of investment which--

ROBERT TOWNSEND: There really isn't any realistic human capital in this model. These guys are choosing whether to be-- set up a firm or have-- or what occupation. You could imagine a model where the choice is whether or not to go to school, but it's just not here. And it's not even realistic in the sense that those x's, you might say, well I went to school, I'm pretty talented, and I know how to run a business, might want to do that. But that talent can change from one period to the next, so it's not great for human capital.

OK. So here are income, savings, labor share, the fraction of entrepreneurs, and a measure of inequality-- the Gini-- as both predicted by the model and what is actually in the data. The data are up here, actually. So this didn't work too well. Income growth in the model is always less than it is in the actual data.

Remember my little bit for science? Where I said let's see how well a model does? Yeah, we can match these things much better if we also choose the parameters to try to match them, but we didn't do that, we're completely under-- it's amazing, it's doing in some sense, as well as it does at all-- like look at inequality, for example. You can't probably tell from the back row the difference between the model and the data.

There's some intervals here. Basically, when you estimate things, you have standard errors on the parameter estimate. So there's a whole confidence interval around all those curves I'm drawing on the occupation choice. So we generated lots of different paths drawing parameters within standard confidence intervals.

But the actual mean path is right here and the actual Thai data is right here. So this inequality kind of is basically very well-tracked, and eventually goes down. When does this start going down? That's when there are no more farms. Well that's kind of the model version of it. There's a sharp increase in the wage at around 1992 in Thailand. I think-- I don't know when the turning point was in China. But this is pretty typical. Fraction of entrepreneurs we're underpredicting, although eventually it rises. Labor share would do quite well, and so on.

So now with the model with its pluses and minuses, we can do some welfare experiments. And what I want to do essentially is a couple of experiments. RCTs. One partial equilibrium and one general equilibrium. So here's my randomized control trial for partial equilibrium.

I'm going to take a household and move it from the non-intermediated it sector and put it in the intermediated sector and then look at the welfare gains. Now actually, before 1992, the wage in Thailand is the same, it's still at the subsistence level regardless of this expanding financial sector. So it's both a partial and a general equilibrium model, but I don't have to worry about the general equilibrium effects of the wage. It's not like I'm giving the economy intermediation or not. Or it's equivalent with having a guy moved from autarky to financial intermediation, because a wage is the same.

So what the hell is this? This huge spike here are basically the gains equivalent-- consumption gains or wealth gains for households of low wealth but high talent. So I pinpointed them earlier in the occupation choice diagram. So they are no longer constrained. So there are some very talented people who will now set up businesses and even hire workers and make a lot of money.

After all, the flip side of the wage being low is that when you're a firm, you don't have to pay high costs. You make a ton of money. So these gains are so big that they make the other things small in comparison. Actually, these sort of orangey people are switching from workers to firms, and these blue-- light-blue people, those were-- I don't know if this will surprise you, those are inefficient firms going out of business.

I mean, they're an autarky, not much to do with their money, throw it back into the business, drive the marginal return down to the carrying costs of grain, and that's the best thing other than being a farmer that you can do. But when the interest rate goes up early on-- and early on the interest rate in this economy will be high, because there isn't that much wealth, and there's relatively high marginal product in production.

So these lazy, untalented, richer people-- sorry, lazy was excessive. They don't have a choice about it, they're just not good at anything. They take their money and they put it in the bank and they get a higher return. We'll come back to this later in the class when we actually come back to RCTs and IV and all of that, but there is already a caution here-- some things are not monotonic. The treatment effect is not monotonic in this model.

Intermediation it's a good thing in the sense that everyone can earn higher income at least when there are no general equilibrium effects. But the way in which you earn higher income may involve occupation shifts-- some people going into business and some people going out of business. So in a very simple way, you can see how you would lose monotonicity of the treatment. We'll come back to that in a later day.

So this is a prediction-- oh, now let's jump countries. I'll go to Mexico. Actually, I think the next slide is-- oh, how annoying. OK, I'm learning. We won't have this-- I'll try to avoid this problem in the future. In Mexico, the fractions of basically wage earners, the fraction of agriculturalists or subsistence people, and the fraction of entrepreneurs. And the earlier one I didn't show you on that scale actually shows these can be slightly different, but on this scale with all the occupations, it works quite well. So yes?

AUDIENCE: [INAUDIBLE] imagine the fractions in each occupation. You mentioned some of that-- there's a lot of movement in cross-occupation type. That probably is a [INAUDIBLE] from the data.

ROBERT TOWNSEND: Yeah, we're not--

AUDIENCE: Times the time.

ROBERT TOWNSEND: Right. We're not actually looking at the panel of a given household at their career choice or occupation choice over many time periods. We're just only using the cross-section to do the estimation. So even though-- I think your point is-- even though this seems to fit well, beneath this picture, there may be households flipping from one to the other and the model doesn't say they should.

Actually, when we get to Buera and Shin and so on, that will be at center stage. They're going to be more realistic about the talent distribution, and they're going to even have forward-looking dynamics so you'll-- they'll anticipate the way in which their talent may evolve in the future and it will remedy this shortcoming.

Now it looks like I'm claiming a big success here in the sense of jumping New Mexico and all of a sudden showing you the outcome after all the same steps, but let me show you something else. So this line is the model predicted number of entrepreneurs in the sector without credit, and this is the one with credit. It's bad news. It's kind of the opposite of what you think. Financial autarky turns out to be better in this sense.

Now normal people don't show you this stuff. But I'm showing it to you for a reason. And it's in that book that I mentioned, the one we're writing on Mexico. So what went wrong? Well I went back and I've been worried about this for quite a long time. I went back and looked at the chapters again. Am I going to put this book in the trash can?

And what I realized was that occupation choice where you do these probits of-- the variables like credit weren't even showing up there as being significant, whereas they do in Thailand. So the data-- the micro data were essentially speaking to me and I wasn't hearing it for a while. It makes the point of the lecture, which is you've got to take the micro underpinning seriously.

And I followed an algorithm as if I'd figured something out in Thailand and I could do it in Mexico. But now I know why it wasn't working so well in Mexico. It doesn't tell you what the next step is, by the way, but clearly LEB isn't going to work there very well.

AUDIENCE: But isn't the firm bigger In the credit sector?

ROBERT TOWNSEND: In the model?

AUDIENCE: Yeah. The firm is--

ROBERT TOWNSEND: Yeah, the model doesn't work.

AUDIENCE: No, but-- so if there is a lower percentage [INAUDIBLE] but the firm's bigger, that's fine, right?

ROBERT TOWNSEND: Oh. All right, I don't have firm size. I'll go back and look if we actually-- oh, I should say also something related. We have-- like I have for Thailand, and it's available to you on Dataverse, we have a massive archive of Mexican data. It should be on Dataverse, and I meant to include-- I hope there is included already Jorge Moreno's summary of the Mexican data. And this book will kind of give you a sense of the great variety of data we have and-- I can go back and look and see whether we actually plotted firm size.

OK, another model. So this is the one with endogenous financial deepening, and I'm leading off with a picture that should remind you of the lecture of last time, which is this trade-off between-- potentially between growth and levels and also the variability of growth. In this case, at low levels of income, you have higher variability, which is kind of a fact that seems to be true in the data. However, this is a model-generated path in which over time, per-capita income is going to be changing, and from those points on, is going to change the statistics of the model in terms of expected future growth and the variance of the growth.

So how does this model work? Well it is-- this one is forward-looking. So households don't just care about today and tomorrow, they care about the discounted expected future, and they care about, in this case, the log of consumption at every date. In a minute we're going to talk about some-- only slightly more general power functions. Beta is the discount rate, and we're going to simplify sectors and occupation choices, almost like telling the usual riskless and risky asset.

Rather than thinking of them as occupations, there's a riskless asset. So when you invest at t minus 1 a certain amount, you get delta times that amount back in the following period. A constant rate of return, and the risky technology when you invest at the amount i at t minus 1, you get it back plus or minus 2 components. There's an aggregate shock, theta t, no i on it. It's common across every household. It's like aggregate risk. And there's an epsilon, which does have-- this is household j, sorry, not i. i as investment, j is household. This epsilon j is the idiosyncratic shock that's hitting household j, and definitely different across different households.

So if you're lucky enough or you've paid the money to be in the club of the financial intermediary, there's an advantage-- two advantages, actually. Information and pooling of idiosyncratic risk. So the return you get at t minus 1 for carrying stuff over is the maximum, largely, of either the safe return or the risky return. But the risky return isn't risky anymore. This model goes too far, you see it perfectly in advance, so you can kind of pick off what to do. It's an extreme version of intermediaries have an information advantage and they can advise clients because they have that advantage.

Well not quite, there's kind of an intermediation cost and that's gamma. So this is a wedge, this is in inefficiency-- it's not modeled in a deep way, but it uses up resources. And we'll come back to that momentarily. And the other thing is, it's costly to get into the investment club, and it cost basically alpha. Not the alpha of the previous slide, just some new number.

There's perfect competition among financial intermediaries, they might be tempted to try to make profits, but they're going to be driven down to basically making people pay the fixed cost to enter the financial system at the true cost and to give them a return that exactly covers the wedge and is otherwise equal to the real rate of return on the investment.

So where did the epsilon go? I said they got the risk-- I didn't quite say it right. You get the risky thing for sure, not only because theta, but because these epsilons disappear. And the epsilons are pooling across many, many, many households. It's IID, essentially, so it goes to 0 as you have more and more households. So it's like a mutual fund. It's not just a bank with fixed borrowing, it's more like equity. You put your money in the bank, it gets kind of the average return on the bank's portfolio, you might want to borrow some to finance investment, but that's kind of separate.

So this looks like a discrete choice problem. And if you bear with me, It's not as bad as it may seem. Suppose at day t, you've already made the decision to keep yourself out of the financial system. So then you have a value function that depends on your wealth, which is not b here, but k at day t, and you're going to decide what to do with that. You can save some of it, and the difference between what you had and what you save, you eat. And the amount that you save, you get to put into-- some fraction into the risky thing and some fraction into the safe thing.

Now you're not in the bank by the decision here, so you're definitely going to be experiencing this idiosyncratic risk. It may make the save return more attractive, but we'll see. OK, so then the next period rolls around, and then at the very, very beginning, you kind of get to decide whether to join the financial system or not at that date. You're forward-looking.

So the value to, say, not joining the financial system is this W that was over here. W means withdrawn or not participating. But if you do decide to join, you're going to get the other branch, the value of participating, and that's basically going to be something called V. However, not quite V, because that instant that you join, you're going to have to subtract off those fixed costs q.

And in this model, it's once and never more, actually. Once you pay to be in, you stay in forever. So this could have been a simpler slide in the sense that once you join and you join the V branch, you'll never again have to choose to withdraw, you'll just stay in the financial system-- you reap all the advantages of information and risk pooling and there's no reason to leave. You'd have to change this model to get disintermediation. But the focus here was on increasing intermediation.

So this is probably something you've seen-- I'm not quite so sure what you've had in your other courses. This is a dynamic discrete choice model. And basically, you have this value function W for low wealth. It's the highest curve, and then you say, no, well this is above it. No, not really, because at the instant you join, you have to subtract q.

So you take this V and subtract q off of the capital level that brings the whole curve down. So where these-- to this. So where this thing crosses, that's kind of the key entry point. At k at 15, given the way I've drawn these curves, to the left of 15, you won't be participating, and to the right a 15, you will be.

Now remember, there's not a whole lot of heterogeneity either, sort of wealth is the key unique state variable at the beginning of the period, so wealth alone is going to predict if the model is accurate, whether or not you're in the financial system, and I'll come back to that. Here, here are these savings functions and portfolio choices, and 15-- or I guess I said this is kind of the key mark. One thing that makes sense is the dynamics, which is you're saving up more and more in advance for this eventual entry into the financial system.

Exactly when you will go in depends on the sequence of shocks that you experience, and you may experience setbacks that push your wealth down even though ex ante your investment decision made sense. But kind of slowly over time, there's this pressure moving off to the right. So it makes-- but there's discounting, right? So you don't want to save up too much early on, you kind of want to be close, and then you don't want to take all that fixed cost out of your wealth at that date, you save up to finance it.

This is unexpected ex ante. So the closer you get, the riskier you go. So you would think people who can't diversify outside the financial system would be doing the safe thing. But actually, this fixed cost is creating a non-convexity. If you remember way back when you see Friedman and Savage, sort of that utility function which is not concave everywhere, well you've got two branches. You've got this value function for participating and the value function for not participating. And in principle, they kind of cross like a butterfly, right?

So then you're getting this non-concave part in the middle, and that's creating this kind of risk behavior where they start doing riskier or risky things that kind of span-- take a chance of grabbing the brass ring and bootstrapping up to a higher level. We didn't see that coming. Actually, those diagrams look like they paced smoothly into one another in those value functions. That's after this randomization is going on. Otherwise-- and so the butterfly kind of disappears, but only after this randomization.

So let me just say a few more in terms of the actual functional forms. Constant relative risk aversion of essentially degree sigma. Beta, I said, was the discount rate. Sorry, this is now xi rather than theta. But in any event, we have to specify-- and they're uniformly distributed-- again, not my favorite choice anymore. If something is uniformly distributed, you have to have an upper and lower bound. So we have picked up two more parameters for the epsilon distribution, and even something for the risky distribution.

So discrete choice-- d for, I guess, discrete choice, basically you're going to enter or not enter depending on whether your value of participating is higher than not participating and vice versa. The model that those parameters is going to give you the k star, which, again in the example, was 15.

In general, k star is going to depend on this-- all these list of parameters, and this list of parameters-- the fixed cost, the discount rate, aversion and utility, the wedge in intermediation, the upper and lower bounds on those distributions, all of those things-- theta for GI-- basically determine k star. And k star is going to determine whether you at a given wealth choose to enter or not enter.

So I'll skip this, but basically there's a complicated way in which the parameters of the model through the lens of the model are generating this sort of threshold value. And likewise, therefore, you can imagine, if at a given wealth, you kind of see the fractions participating and not participating in the cross-section, that you would be able to sort of invert things and estimate these parameters.

It's a bit tricky because that theta, the level of the aggregate shock kind of matters, so we have to integrate up over that, too. But when we do the estimation using a socioeconomic survey, the papers online, we ended up estimating-- and actually, comfortingly, a lot of the parameters make sense. The discount rate's close to 1, but not exactly. A degree of risk aversion is like almost the log case and the depreciation rate was 4% and so on and so forth, quite remarkable. Yes?

AUDIENCE: And so is that-- am I reading this correctly? Is that gamma was estimated to be 1?

ROBERT TOWNSEND: In this case, yes.

AUDIENCE: OK. So with that imply efficient-- like there's no wedge?

ROBERT TOWNSEND: Yeah, no wedge. Yeah, that-- I'll come back to that momentarily. So again, we can simulate the model-- and under various estimated or robustness checks, parameter values-- to get the growth rate, the participation rate, and the tile index of inequality.

Now one caution on the growth rate, and I always have to check to make sure-- this model has aggregate shocks. So if you integrate out over the expected shocks, you're going to get a smoother growth rate path than in the actual data. And what is it that I have to check? I always have to go back and see whether this was one where we use simulated values of theta or we took the average over those thetas.

We missed this upturn here and the growth rate, and that's related to the fact that financial participation goes flat right here. This model's great at the trends. It gives you this very smooth, ever-increasing participation in the financial system, and actually, it gives you for a while a good shot at the increasing inequality as those firms made a lot more money, but it misses the downturn.

In fact, there are no real prices, there's nothing like a wage or an interest rate, that's a defect. And so it's almost like a linear world, in some sense, and we can't get inequality to go down in this model, but we can get the trends right for a while. This participation thing is going to bring me back to the wedge.

So what do we do? We're going to do a policy experiment, basically. We look at the government's control of the banking system in terms of the ratio of credit. It's moving around a bit, but essentially it's flat, it goes up, and then it comes down again. This is the liberalization and this is the retrenchment that happened before it.

So what we do is, we vary that parameter gamma. It's not going to be costless intermediation. In fact, we're going to subtract off-- I think it was something like 3% or 4% when the intermediation is being done by government banks. And we don't change the model otherwise.

And amazingly, we get this almost perfectly right. So we can actually get the slowdown in the economy as a function of this, quote, bad policy experiment where the government, for one reason or another, essentially took over the banking system. By the way, we don't do very well with growth differences even after liberalization, and we can-- and the reason is partly because, at least through the lens of OLS regressions, it's-- I've alluded to this before-- it's just basically junk. I mean, it's not a stationary model.

The decision rules are stationary, but aggregates like growth rates are not. And you can use this value functions to back out the welfare gains. There are a number of ways to do that. One thing we could do is just surprise people. They think the banking system is going to stay restricted, and they've made their dynamic plans, and then we surprise them and liberalize. And they're better off, and we can calculate through the value functions. It's a bit complicated and I'm not going to take you through these slides, but the welfare gains can be quite large.

Now we can imagine surprising them, like varying that wedge or varying the fixed cost and it actually matters a bit, and we can take this model to Mexico and it does pretty well with the trend of the growth rate, but around 1992, Mexico had yet another crisis. And not surprising, this model is just very ill-equipped to handle that kind of stuff.

Actually, things started going better with Vicente Fox around 2002. We could restart the model over here and it will do well. But it's not going to explain-- this is why Mexico is still poor. I mean, they go upstairs, they go upstairs, and then they fall down again. And they've been doing this since the early '80s, you know? Now hopefully they're not going to fall down again.

But the model-- I promised you shortcomings, right? This model is not designed to get at sudden stops or things along that line. So the last thing I want to show you quickly is another kind of experiment. Instead of going to a new country like Mexico, we started to look interior to the country of Thailand.

This is a picture of basically wealth using a principal component. You can see in and around Bangkok and going north to Chiang Mai-- the wealth corridor, if you want, of the country, these green things are the provinces of my survey. And looking at them again, you can actually see where the villages are, you can see the road network, at least on major highways, you can see the district centers.

And I promised I would ask you this. I mentioned it before. If you have an interest in going through some of the GIS capabilities, I'm more than happy to schedule an extra session. I have to check my GIS person. I see some of you shaking your heads. You don't all have to go, so that's enough for me. So let me check with her and maybe we can possibly do that Friday of next week, but it would be over and above the regular recitation section.

And so here is a picture of where the businesses are in Thailand. And through this GIS, something called the Moran, you can see whether areas, which have high levels of enterprise, are surrounded by other areas with high levels of enterprise or low levels surrounded by low. And so you see this quite salient geographic picture.

And here, we took levels of enterprise in 1986 and looked at the growth rate of enterprise between 1986 and 1996. This area in here, the colors are a bit pale-- it doesn't come out too well-- are areas of convergence in the sense that they started at high levels but then they shrink to lower levels.

This area around it, though, these reds, those are areas of agglomeration. They have high levels and high growth rates. There's an ever-increasing concentration of enterprise. These pink areas are catch-up areas along the corridor which had low levels, and now they're gaining a lot of enterprise. And these blue areas out here are kind of like the boonies. They have low levels and low growth rates.

So it's not a uniform picture. In fact, you can take this down to the provinces and look in the data at the areas of agglomeration and relative low levels, and then that looks like almost the same picture. That's a model-generated picture. So we actually projected-- took Lloyd-Ellis and Bernhardt and projected it down at the village level, and here you can see agglomeration. It's not exactly identical. There are certain predictions in the model that don't happen in the data, but largely these sort of favorable areas-- a subset of the candidate of favorable areas are the ones that actually start growing like crazy in Thailand.

And here at that same level is Greenwood in Jovanovic. And here, these red areas are areas where the model overpredicts the data, and the green areas are places where the model is underpredicting. In other words, these red areas, which turn out to be the same urban areas, Greenwood and Jovanovic is predicting they should have more banks than they actually do. So in some sense, even though they're higher wealth, they're underserved relative to what the model would predict. And these green areas, those are areas where there's more access to credit than Greenwood and Jovanovic would predict.

Now it turns out this story has to do with the interplay between commercial banks that operate in urban areas and the bank for agriculture which is operating in rural areas, and we'll come back to that in the last lecture of the class. So I'm done, essentially. I'll just point you to the rest of the slides. Basically we did kind of a rigorous comparison of the successes and the failures. So you can actually go through each of the models and see on which data which model's doing well and which data models are doing poorly, which feeds into the issue of model selection and next steps in the research. So OK, with that, I'm slightly two minutes beyond.