



markus<sup>o</sup>academy

## MOTOHIRO YOGO ON FINANCIAL INCLUSION IN THE US

January 27, 2022

### Webinar Transcript:

0:00

Markus Brunnermeier: Okay, welcome back everybody to another webinar organized by Princeton for everybody worldwide. We're very happy to have Motohiro Yogo with us from Princeton, hi Moto.

Motohiro Yogo: Hi Markus.

Markus Brunnermeier: It's great to have you with us and you're presenting some joint work with Natalie Cox and Andrew Whitten. It's titled "Financial Inclusion Across the United States," so we're looking forward to learning more about financial inclusion. And, in particular in the light of fintech and other developments and how racial divide and other aspects can be better understood in the financial inclusion world. Before I pass on the floor to you, let me just give a few opening remarks. The first one, I would like to connect financial inclusion to resilience, and you know if you have better financial inclusion, you might be better to bounce back after you face a shock. And how can you include people better? Once they can save better, they have better saving tools, so that helps them ex-ante to prepare for some shocks and be more resilient down the road. Or they can also prepare and be able to be better insured so that they get some compensation after facing a negative shock that also gives them some flexibility to bounce back and not be trapped in some traps or going downward spirals, and all this, all of this financial inclusion helps and it might also be related to the resilience inequality, I was pointing out in my book and earlier, if you have different people, some are able to bounce back and others are not able to bounce back, even though they have the same income and same wealth. The former which are able to bounce back, they actually have a higher resilience and hence they can take on more risk, they can earn some risk premia, they can take chances and opportunities others can't and that might down the road lead to some wealth inequality. So resilience inequality is also very important, but hopefully financial inclusion will reduce this resilience inequality. The other aspect that I would like to raise in terms of fintech: how does financial inclusion look today, and how does it look in the future? So fintech might help us to improve financial inclusion. That's because today the marginal cost of providing these financial services through these new technologies are much lower so that's essentially the good aspect, but let's also look a little bit into the future. So financial inclusion might become essential or necessary, so you mentioned today, the Internet, essentially, people say it's a human right to have Internet, or electricity is a human right of electricity, perhaps in the future it will be a human right to have financial inclusion, because we can essentially not live without financial services, so the power to not be included— actually the punishment is much more severe, so that is actually the fintech companies also empower to exclude. If you throw somebody out, you can punish somebody. So imagine, these days, you can always fall back on cash, but if there's no cash anymore, we do everything with online payment and suddenly the platforms kick you out. The exclusion power is way more powerful once you have financial inclusion. We all got used to it, and I would like to

understand this as well. Of course you could argue that these platforms have better enforcement technology, they can enforce that contract, because they can exclude people and punish them severely. But the question is, do we want that, and so, who should? How should we regulate this to make sure that financial inclusion by some fintech companies doesn't become too powerful and it's still in the private interest, even for poor people. So that's essentially another issue I would like to raise, you know who is deciding, this is private firms deciding it, whether you can exclude somebody from services or not, and then I would like to come to the questions. Moto raised some poll questions and I'm curious to see your answers and here are the answers to the first question "Do you know, an American or dad who does not have a bank account," the question was answered this way: 13 percent said yes, to know somebody, but the majority 87% don't know anybody who is not a bank account. "Do you know anybody who doesn't have a retirement account?" That's much more 50/50, so 54% know somebody and 46% don't know anybody. The third question was "What do you think is the most important factor in the geographical difference which determines whether financial participation occurs or not?" Is it race, that's what 16% said, is it income— 62% said income is the most important factor, and access to banks and other financial services was 20%, and other factors only 2%. And finally, the fourth question was should the government require—the state or federal government—to offer that employers have to offer some retirement accounts to the employees and over a third strongly agree with that and another third agree with that so it's two thirds essentially have agree or strongly agree. 21% are neutral, 8% disagree, and almost nobody strongly disagrees, only 3% strongly disagree. So with these answers I give the floor to Moto who will enlighten us. What he has learned from looking through some fantastic data about financial inclusion in the United States across and what are the driving factors, whether you know people like included or excluded and that can be part of what the financial services industry is providing. Thanks again, Moto, I'm looking forward to your presentation.

5:54

Motohiro Yogo: Thank you very much Marcus for this opportunity to present my work, joint with Andrew Whitman who's at the U.S. Treasury Department, and my colleague at Princeton, Natalie Cox. So, kind of the starting point for this project is the ideal that in an inclusive society, everyone should have access to financial services, in particular banking and retirement accounts, regardless of your income or race. We kind of know from household surveys income and race are important determinants of financial participation. As wonderful as household surveys are, they do have some limitations, one is that they're typically small samples. They have a limited panel dimension, for example, the survey consumer finances, we observe cross sections of households that are not necessarily linked over time. And then the third is potential measurement error that comes with survey data, so in this project, we use big data from essentially all the tax records for U.S. households and we focus on age 50-59 which is kind of the critical part of your life cycle, where households are starting to save about safe think about savings for retirement and we covered the sample from 2008 to 2018. And what the big data allows us to do is do the detailed geographic analysis that would not be possible with survey data that are smaller samples, so we can really look very closely at the geographic zip code level to see how financial participation correlates with things like income, race or access to financial or banking services. And then the second thing we can do is, we can evaluate recent state laws that require employers to offer a retirement plan and I'll be talking more about that a little later. Okay. So the administrative tax data— we actually have data from 1999 through 2018, so 1999 is the first year in which the tax data were digitized. So okay, so that's our starting point. And most of you know about form 10-40 you know those are the annual tax returns that you file.

But in addition, you might get these forms from your employer, such as a W-2 for wages, a 1099-INT which reports interest income from a financial institution, a 1099-DIV (dividend), a

1099-R which are retirement accounts, and so these information returns allow us to cover a population that's much wider than the population that's actually filing taxes, okay. So even if you don't file taxes, you might actually receive a W-2 or SSA-1099 from the social security administration. So by these information returns, together with tax returns, we can cover almost the entire US population, so we get a coverage rate of about 96% relative to the population census. And we can tell which individuals are married through joint tax filing or even if you don't file taxes together this particular year, we can observe that you filed in the past together. And that you live in the same address, so through those data, through those information, we can link individuals together and form households. In our measure of income in the data is we observe a five year history of your income, we take an average. And we call that "usual income." So okay, so this five year averaging allows us to kind of say what your usual level of income is so smooth it over, like bonuses or fluctuations that are only temporary.

10:02

Markus Brunnermeier: So, Moto, can I ask you a question? You know a lot of the unbanked people in the U.S., you know, some of them are illegally here or they are in hiding, so does your data capture them too? Essentially only the guys who filed taxes, but if I'm illegally here do I have to file taxes. Do I file taxes or not?

Motohiro Yogo: Oh so just to be clear, the information returns go out to people who don't even file taxes. Right, so if you work for an employer and earn wages, I get a W-2 now whether I file a tax return is my decision and lots of individuals choose not to file, even though they receive like W-2, because their income is below the threshold for paying taxes. Now some of these individuals maybe should have filed taxes to receive a refund but not everyone does file taxes. So our coverage is much wider than just tax filers, for that reason. And then the other part of your question is what happens with people, for example, who don't have social security numbers, those are also in our data and they're assigned in a data these identifiers that allows us to identify who they are, but they don't necessarily have to have a social security number. Now what they do have to have is that they have to be issued one of these forms, so again, you might be issued a W-2 or 1099-MISC for miscellaneous income, and these are forms that you could receive even without a social security number.

Markus Brunnermeier: Do you have any idea of what fraction of the population living, the U.S. is not covered is 1% or is tiny?

Motohiro Yogo: So 4%, yeah so we're missing 4% now. I should mention that we're actually in the process of improving the coverage further. So we recently discovered that some people don't receive any of these information forms that I mentioned, but they do receive something like a 1095, which is the new information return that came along with the affordable care act, which tells you whether you're covered by medicaid, for instance, so it turns out that once we include those we're going to get much closer to 100, I think we can get essentially 100%. So, but that's kind of in current— under works right now and we haven't implemented that yet. Okay, so how we measure bank account participation is based on the 1040, the tax returns, so if you receive a refund or make a tax payment using electronic funds transfer and that gets recorded in our data, we observed that you have a bank account. The 1099-INT also tells us whether you've received interest income. Now, the problem with 1099-INT is there's a minimum threshold, you have to earn interest income at least \$10 for the financial institution to be required to file a 1099-INT on your behalf, so this doesn't isn't always a reliable source of information for whether you have a bank account. For retirement accounts, the measurement is, I would say even better in the following sense, so for all of us who receive a W-2, there's a checkbox on W-2 that says

whether you participated in an employer retirement plan, and that includes both defined benefit plans, as well as defined contribution. A 1099-R is a distribution from a retirement account that includes both employer accounts, as well as it includes IRAs, individual retirement arrangements. And then, a 5498 is a form that gets issued for contributions into an IRA, so by combining these information returns we basically observe every retirement account and transaction, including defined benefit plans, defined contribution plans, as well as IRAs. Now, in addition, we also want to measure whether you have access to a retirement plan, through your employer. Okay, so not all of us whose employer offers us a plan would necessarily decide to opt in. Some people decide to opt out, so we want to see whether people are opting in or opting out and the way we can measure that is again based on W-2s. So I observe all the W-2s, right, so I can tell whether for a given employer and so EIN sensor, employer identifier number so for a given employer, I can see whether the checkbox on the W-2 is checked to see if anyone at this firm has a retirement account, right. So from that, I can tell whether the employees and not employer are eligible or potentially eligible for contributing to retirement plan. Now for each of these measures, we look at the panel dimension, so we have a 10 year history of whether you've ever received a refund or tax payment in the past. Or whether you've ever participated in an employee retirement plan from your history of W-2s, so once we consider the fact that we need a 10 year history to measure these variables, we end up in their final sample of 2008 - 2018. Okay, so one thing we are concerned about is maybe because it's actually...

15:47

Markus Brunnermeier: So you threw away some data because earlier you said from 1999 onwards, now you want to start in 2008 is correct?

Motohiro Yogo: Oh yes, because the 2008 observation that I have is based on the history from 1999 to 2008.

Markus Brunnermeier: Okay.

Motohiro Yogo: So for each year, I have a 10 year look back period. Yeah, so that's 2008, the earliest year in which I have a full 10 year history to measure my variables. So one thing we were worried about is because the tax returns are kind of a flow based measure right that's like whether you contributed to a retirement account this year, or whether you filed taxes, this year, using electronic funds transfer. There's a potential risk that we under-measure participation rate, because you could be in a situation where you have a bank account, but you happen to not file taxes, this year. Okay, so one way to kind of assess whether we're severely under measuring financial participation is, we can take the average of financial participation in our data so in 2016 households at the lowest income quintile, so this is a lowest 20th percentile of income, 82% of those households actually had a bank account, according to our measurement. Now if we're under-measuring this, then this should be lower than what it is in the survey consumer finances, which is actually 76%. So, if anything, we're actually getting a higher estimate of financial participation, then, then the SCF. Now, if you look at the highest income groups, okay so households in the highest income quintile between 80 and 100%, then 100% of those households have a bank account, regardless of whether you're measured in the SCF or in the tax data, you know. Okay, so kind of the measurement at the top is not affected by which data you use, but the measurement at the bottom seems somewhat sensitive to which source you're looking at, whether the text data...

Markus Brunnermeier: So what does the SCF data have that the administrative text data doesn't. In a sense, you said it almost has 96% of the population, no?

18:13

Motohiro Yogo: Yeah so there's two hypotheses, one is that the remaining 4% that we don't measure might be all non-participants. And that would reconcile this gap and so that we're trying to close in on that hypothesis, by trying to get closer to 100% but, but the other one is SCF is not a perfect sample either. Right, so it is a survey, it does require people to, you know, answer the survey and answer them accurately and so on, so kind of the way in which economics research is proceeding these days is survey data are still very useful and we still use them. But administrative data gives us an additional measurement where sometimes we discover that things that we thought were facts in the survey data are no longer true...

Markus Brunnermeier: So the survey of consumer finance is— are they zooming in on the very poor ones or they just...

Motohiro Yogo: It's supposed to be a representative sample of the U.S. So in theory we're trying to measure the same thing but in practice it's not exactly the same.

Markus Brunnermeier: For the standard errors on this 76% could be quite high.

Motohiro Yogo: So I computed the standard errors, the standard errors are no more than 1%.

Markus Brunnermeier: Okay.

Motohiro Yogo: So 76 plus or minus 1% is the standard error. So it's very, it's pretty tight actually so the standard errors wouldn't be able to explain the differences.

Markus Brunnermeier: How many people are involved in the survey?

Motohiro Yogo: Around 7,000, in our age group 50 to 59.

Markus Brunnermeier: And then your tax data, there are millions.

20:07

Motohiro Yogo: yeah in my tax data I have 6 oh Sorry, I want to say six, over 6 million, it's more like 7 million. Yeah because I literally measure every single person in that age group yeah. Okay, so in terms of retirement account participation these numbers actually are pretty surprising and in some sense. So one thing that's not surprising is, if you look at the highest income group so 80 to 100 percentile. Essentially 99% have a retirement income so retirement account participation is not a problem for the wealthy. Maybe they find a way to participate. But even middle income participation is still pretty good as I'm circling here. 98% or 97% in 2016 in the 60 to 80 income group, still 91% in the 40 to 60 income group, but then it kind of drops off so among the lowest income group, it's only 42%. So this is what I mean by financial inclusion is why, why is it 42%, why is it not higher? Is there anything that society can do to make this 42% higher? Those are the kinds of questions.

Markus Brunnermeier: What's puzzling to me is that the SCF, the survey of consumer finance, is always lower than the tax data for any income.

Motohiro Yogo: Yeah, so this predates our study, but some people have also looked at the difference between the administrative tax data and the census survey data and what they

always find this survey data systematically seems to underreport retirement account participation relative to...

Markus Brunnermeier: Also people don't want to report that they have a retirement account.

Motohiro Yogo: It's either that they don't want to report, or they forget that they have one so suppose like you're not a consistent contributor, or you switch jobs in the last 10 years you might have contributed, five years ago, at your previous job, but you forgot to report in a survey or if you haven't defined so if you have a defined benefit plan it might be even worse, because if you have a defined benefit plan you're technically might have a retirement account but you're not contributing to it actively so you might not even remember realizing you have one.

Markus Brunnermeier: So I should trust your data more?

Motohiro Yogo: Yes, subject to some caveats but yeah we're trying to literally measure every single person so that's kind of the spirit of the exercise. One thing that's interesting is how are people participating in retirement accounts, so the key numbers in the lowest income group of 42% participate right, but then is it through employers, IRAs, or both? So 25% only have an employer plan, 6% have only an IRA, 11% have both. So predominantly among the lowest income group, the way in which they participate is not by going to fidelity and opening up your own IRA. They have a retirement account through their employers, so the combination of 25 and 11, that's 36 out of 42, yeah, most of them do participate through employers now if we go to the highest income individuals or households, these are 99% of households in the highest income group participate among those households, 76% have both. So kind of predominantly wealthy people find a way to participate, not only through the employer plans, but also through IRAs. Now I should add a caveat here that having an IRA doesn't necessarily mean that you contributed directly to an IRA, if you have an employer account and you quit your job, you can roll over your employer plan to an IRA so some of these guys have an IRA, not because they've contributed directly to an IRA but potentially through through their employer plans that they roll over.

24:44

Markus Brunnermeier: So, Moto, there are some questions from Tom. He would like to know if the IS knockout data would enable you to identify specific taxpayers and that might lead to some change in the percentages.

Motohiro Yogo: What do you mean by specific taxpayers, I'm not sure I understand.

Markus Brunnermeier: If you could identify a particular taxpayer, that would remove them from your data set, probably not, no, because your percentage points...

Motohiro Yogo: Well what do you mean by removed? Sorry I'm still not...

Markus Brunnermeier: Because they want to make sure that you cannot identify somebody I guess they only remove really very wealthy guys from your tax data.

Motohiro Yogo: Okay, now I understand the question. I'm sorry, but okay, so first, the data are anonymized in a way that makes the identifiers are not actual social security numbers, so they assign a kind of pseudo identifier that doesn't allow me to identify people. And then the second one, this is kind of critical and this, the most important part is that I'm not using dollar amounts

of contributions, I'm using one or zero of whether they have an account or not, and that actually allowed us to do this study, because if I were using actual data on dollar amounts, that would obviously be more sensitive, because that might allow me to identify very wealthy individuals.

Markus Brunnermeier: So, essentially, if I open an IRA account and put a hundred dollars in and I forget about it for the next 30 years, I'll always have an IRA account.

Motohiro Yogo: Yeah exactly. And in fact, in your example that you just said, the institution, where you have an IRA account would continue to issue these 5498s. So I would observe, you know in our data. Okay, so one of the kinds of the facts that popped out of our study, which is kind of interesting, is what's happening to financial participation over time, so like our kind of prior before looking at the data was "well things have to improve over time, right?" So that turns out to be not so true, so what this figure is showing is that if you're in the highest income group, they're always participating, one, right. And then, if you're lower income, well you're nearly almost always participating. But there's no meaningful trends in the data now. If you look at the lowest income group then among these households a participation rate in 2008 was 85%, and this gradually declines over time to 79% in 2018 now at each year what I'm looking at is the participation rate among 50 to 59 year olds. So what this is saying is that 50 to 59 year olds in 2008 participated more than 50 to 59 year olds in 2018. So these are essentially what we call cohort effects right, each younger cohort of households are participating with less frequency over time. You can do the same thing with retirement accounts, where the trends are not only visible among the lowest income, but even the second lowest income group. Among the lowest income group the participation rate in 2008 was 49% and it's the client to 41% in 2018. Now, I think that the one for banking was more surprising to me than the one for retirement accounts was less surprising to me for the following reason. There are all these narratives going around that the types of jobs people have these days are not as secure as they used to be so temporary workers, gig workers, and so on, they might not have the type of employment arrangement that allows them to qualify for retirement account through their employer and, moreover, we know that the secular trend in the economy is that both private and government pension plans are disappearing over time, private pension plans, because firms are no longer interested in bearing the risk of a defined benefit plan. And then governments because of pressure to their fiscal capacity and so on, so that those kinds of very important trends in the economy could very well translate into these patterns of a declining access to retirement accounts. And then panel C makes this point more clearly, what panel C shows is not whether you participate in the account, but whether you have had access to an employee retirement plan in the last 10 years. So in 2008, 60% of households in the lowest income quintile had access to retirement plans, but now that number has declined to 50% in 2018 so it's actually declining at a rate of about one percentage point per year. So this is the sense in which there's a policy intervention, potentially, which is, can we stop this trend from continuing, and can we reverse this trend, maybe increase access to retirement plans. And as I'll discuss later in my presentation, that's kind of the spirit of the state laws that have been passed across 10 states now, that essentially make access to retirement plans mandatory for all employers. Okay, so to get our feet warmed up for the next part of the paper, where we do geographic analysis and want to remind people like what we see in the SCF regarding bank and retirement participation and how that correlates with things like income and race. So in column one, this is a linear regression of bank account participation, one or zero on whether you're Hispanic, Black, other non-white and the omitted category being white here, and what this shows is that Hispanic households are 14 percentage points less likely to have a bank account than a white household black households are 13 percentage points less likely to have a bank account. If we control for income here, right income is a very important determinant of financial participation, but bank account participation still correlates with whether the household's Hispanic or Black by nine

percentage points difference. So what this is saying is income is important but doesn't seem to explain away the correlation between financial participation and race. And then same thing with the retirement accounts, so Hispanic households are 33 percentage points less likely to have a retirement account, Black households 22 relative to White households. And if we control for income in column 4, these numbers decline a little bit, but it's still a 21 percentage point difference between Hispanics and whites and nine percentage points between Blacks and whites.

32:25

Markus Brunnermeier: So, Moto, can you also make a time frame on that? It is like the situation for Hispanics is worse for the whites, let's say, but over time, the gap is closing or is widening?

Motohiro Yogo: Well that's a good question, so this data that I'm looking at is not sure if it's a long enough sample to do that, so what I should do is go back in the SCF and test that, so the answer is I don't know.

Markus Brunnermeier: So if you look at, for example, life expectancy, you know the situation where, especially for the whites, it's worsening over time, so perhaps you find a similar thing here as well?

Motohiro Yogo: Yeah that's exactly right, so this is just a snapshot in 2016. So yeah, I'd have to repeat this progression over time to answer your question, so I'm not sure. Okay, so based on this type of analysis which is not new to our paper people have kind of looked at these regressions over many studies, there is a puzzle in the literature like, why does race seem to matter, even if you control for income? Is there something special about race, what is it capturing? Is it evidence for discrimination and so on? That's kind of the narrative and the questions people are pondering about. Okay, so given that type of fact we go into our geographic analysis, so what we do is, we can tabulate participation rates by zip code so it's not exactly zip code, but they're called zip code tabulation areas so they're essentially geographic representation so zip codes that are developed by the census bureau and we can link the data to actual population counts as well as the ratio a population shares by zip code and we can also see the location of bank branches by zip code using the FDIC data.

Markus Brunnermeier: Moto, can I ask a different question? Is there any way to identify who offers this retirement account? Is it a traditional provider or is it some fintech companies? Is there any way to see developments? New through fintech, improving or worsening things.

Motohiro Yogo: Yeah, that's a great question. Because our bank account participation measure comes primarily from tax returns, I don't think we can tell.

Markus Brunnermeier: You don't know which bank is filing these tax returns.

Motohiro Yogo: Yeah, I have to look into it, so there is a payment code associated with each tax filing and the question is whether the payment code reveals the identity of the financial institution and whether we can use that data to identify the identity. So the question is whether, whether we have that data and whether we can use it. Those are two separate questions and I'm not sure if I can get back to you on that. Sorry, my slides kind of went blank so let me just try to get them back now. Okay, so this is the geographic map that we created where red indicates low participation in bank accounts and yellow indicates high participation and the areas are zip codes, and this is for the 0-20th percentile in income, so what you see here is it's not like the



South is all red and the coast are all yellow, it's kind of mixed. So, based on this kind of analysis and with as well as regression analysis, we actually figured out that the geographic differences in participation is very significant, even once you control for commuter zones, so commuter zones are representations of counties that feed into the same Labor markets, they're called commuting zone, so there's about fewer than 1000 commuting zones in the U.S. So once we control for commuting zones, we can see that even within commuting zones, there's a lot of geographic variation and participation rates, so what that prompted us to do...

Markus Brunnermeier: Moto, I have a question, I might have missed something, but it looks like in areas with higher population density there's fewer people being banked compared to populations, where those areas where the population density is much lower.

Motohiro Yogo: Yeah it's just— yeah so that's exactly one of the things I tried to see if participation correlates with the population density. It does, but not very strongly, so your eyes are like misleading you....

Markus Brunnermeier: Because if you look at the states on the West Coast, not California, it looks pretty yellow, not too dark. It's just because there's nobody living there.

Motohiro Yogo: Yeah so the shade indicates density to the right, so what you want to look for is very deep red shades that means highly populated but also low participation. And then the point is that that's happening all throughout the coasts and it is not particular to South Coast commuting zones and so on it's like this is like variations happening, even within commuting zones is what...

Markus Brunnermeier: Just so I understand this, the darkness is a population density and the color is the participation.

Motohiro Yogo: That's exactly right.

Markus Brunnermeier: But you know dark orange might be very close to red.

Motohiro Yogo: Yeah so what you're looking for is like a troubled area is something like a very deep red right because that deep red means— red means low participation, deep means it's high population. Yeah, and then, and then you shouldn't pay too much attention to yellow, that's not very— like these areas, right, where it's yellow, but these are low populated. So, I think one way to kind of summarize this data is to kind of run regressions of zip code level of participation on measures at the geographic level, so we can look at racial composition by zip code, average income by zip code, and see also whether correlates with location of banks by zip code. So this is kind of a big table, but let me go through each column and I think it's going to clarify some of the questions you just had. So in the first column, this is like the simplest specification, where we run a cross section of regression of zip code level participation and bank accounts on the fraction of the population that's Hispanic, Black, Asian, other non white and omitted category again is white. So we get about 27%  $R^2$  and you see that these coefficients on Hispanic and Black are negative and highly statistically significant. In zip codes, where the Hispanic population is one percentage point higher, then the participation and bank accounts is one basis point lower. So in column two to what we do is we control for commuting zones fixed effects, so now, what this column two is asking is this correlation between race and participation at the zip code level due to differences in geographic areas that are broad or is it kind of micro variation within local geographic areas? So what the commuting zones do is it says that well there's a

common level of participation let's say in the New York metropolitan area. So once we control, for that is there any remaining variation in the New York metropolitan area that is explained by racial differences. And the answer is yes, so you see that, if anything, the coefficient on Hispanic and Black actually become more negative so now it's like seven percentage point differences okay. So what this prompted us to do as well, so what you really need to be looking for is there something like different within the New York metropolitan area where some households or some neighborhoods that are more Hispanic or some neighborhoods that are more Black, there's some differences in participation rates that's not entirely explained by these geographic measures. So one obvious candidate to us initially was well, maybe it's that in zip codes where Hispanic or black population shares are higher, those are zip codes where banks don't locate. Okay, so there are fewer bank branches in those areas and, as a consequence, these households do not have access to financial services and, unfortunately, that variable seems to do nothing so it's not really correlated with the frequency at which you observe bank branches within the zip code.

42:45

Markus Brunnermeier: It could be very different before all this online banking and office online banking so, can you separate it over time somehow?

Motohiro Yogo: Yeah, so we run this regression like in 2008 versus 2018 which might not be like enough of a time series to tell apart, but it doesn't seem like it's sensitive to yeah it's not a time thing it's kind of a cross sectional.

Markus Brunnermeier: And the explanation is that with online banking the partners don't matter so much anymore?

Motohiro Yogo: So with online banking, yeah, what could happen is that the location of actual physical branches shouldn't matter so this coefficient zero might be explained by the fact that if online banking is kind of the predominant language people think now, then it shouldn't matter right. So you're right that, to the extent that online banking is important, our variable might not be capturing the right thing here.

Markus Brunnermeier: Okay you're not running it for your 1999 data, whatever.

Motohiro Yogo: We could run it for 2008 and we can see that even in 2008 bank branches didn't matter so that kind of goes against that hypothesis, in a sense that online banking might not have been very present back in 2008 and even then you get the same result you know. In column 4, is kind of the magical variable, just average log income in a zip code so once we control for income, you notice that the coefficient on Hispanic and Black goes to zero. So, based on this we conclude that it's not really like racial differences across the codes, it's more like income differences across zip codes that seems to be explaining the local geographic region and participation. And the last column says that different zip codes also have different costs of living, in particular regarding housing, so once we control for the average home value by zip code, we thought we might see some differences, but just didn't seem to matter so much. Right so one conclusion that can come out of this is it's not really something like innately about race, but all that variation we see in participation is really about income, but when we say income, it's not clear what exactly income means so the direct effects of income is obvious ones, like if you don't have enough income you're not going to save, so you have no need for bank account or retirement account. If you have low levels of income also your account balances, even if you open a bank account might be low. Now with low balances, you might be incurring lots of

banking fees, so that might actually discourage low income households from maintaining a bank account. And then, finally, in regard to retirement accounts, one major way in which we recently opened retirement accounts just to have tax efficient savings, but if you don't pay taxes in the first place those incentives might not be there. Now that doesn't actually work exactly because there's what's called a savers credit that allows you to claim tax credits on contributions, even if you're a low income household that doesn't pay taxes, so this argument doesn't exactly work, but if I mentioned it just as a possibility.

Markus Brunnermeier: The fact that income and race is highly correlated, it's taken care of in a sense.

46:38

Motohiro Yogo: Yeah to the multiple regression, we're asking whether even though they're correlated, is it really income or race yeah so we're letting the data speak whether which ones are more important variation even though they're highly correlated. And then other reasons why you get correlation with income might be that income is associated with financial literacy, so it might be that what we're really finding is not that it's not only income that matters, but income tends to be correlated with levels of financial literacy that might discourage some low income households from participating. And finally, there could be some pure effects because we're looking at geographic level variation so it might be that in high income neighborhoods like Princeton, New Jersey, even low income households that live in Princeton are likely to participate, because they're likely to bump into households that are high income and tell you something about retirement accounts. Let me skip over this equation, but then kind of summarize the argument that we make, which is one reason why we might find differences between SCF which is at the individual level and the tax data geography level is that the geographic data are aggregated versions of individuals, so what we're looking at is the sum of what we call individual effects and group effects, or you think of it as potentially pure effects. So what you kind of need to argue is that in order for the differences between the SCF and the tax data to be explained by individual versus group effects, what you need is that the individual effect, which is that blacks are less likely to participate, has to perfectly offset a group effect, which is in neighborhoods with a higher black population share, white households are more likely to participate, so if that's the case, then those two effects can offset each other and in a way that we find these zero effects at this column. But what's kind of hard to argue or hard to believe about this argument is that the individual and group effects have to exactly offset each other to get to zero effect, right, so it's like a knife edge scenario that seems unlikely. And moreover it's not really clear why white households in neighborhoods with a higher black population share are more likely to participate, rather than less, so it's hard to think of stories why that would be the case. So a simpler story that we think is more compelling is simply that income could be mismeasured in the SCF so you get the classic attenuation error in economics, which is in the SCF income is much better it's going to be picked up by something that's correlative income but better measured, mainly race. But once we moderate income more accurately in the tax data, then this attenuation bias goes away and you get a cleaner relationship between income and participation. So that's kind of our preferred story now he's done an analysis of our data. Then the last thing we wanted to show you was that employee retirement plans are an important way in which people participate, and in particular access to employee retirement plans could encourage households to participate, right, because it's a lower cost way to participate than opening an IRA account. Now, the problem is, if you run a regression of whether you participate in your retirement account based on whether we have access to an employer plan, there is an obvious endogeneity bias, which is that workers in the cross section have heterogeneous desire for retirement saving, so those that want to save for retirement sort into those employers that actually offer these items so then that's the bias that

you would get in it in OLS regression. So instead we do instrumental variables and our idea is that we can look at the cross section of individuals in 2018, and we can look back to their W-2 in 1999 when they were much younger between 31 to 40. And the idea is that we're going to condition on the subsample of these individuals who did not have an employee retirement plan in 1999, but they worked in different sectors, okay, so some people worked in let's say finance and insurance and others worked in education back in 1999. Now, because finance and insurance versus education have different prevalence of employee retirement plans, and if you stay in that sector, then over time you're going to be more exposed to retirement plans if you're into sectors that have higher prevalence of retirement plans. So we use the sector of employment in 1999 conditional on whether the employer offered a plan as an instrument for your access to a retirement plan today. So, based on this IV specification, we find that among low income households, for each year of access to an employer plan that increases participation in a plan by nine percentage points, so that's a pretty big effect. And the big effect is also true for middle income households in the 20 to 40 and 40 to 60 percentile of income. Now, based on this IV regression what we do is we do a policy counterfactual which is suppose we increase access to employer plans by offering it universally, and this is not a plan that a policy counterfactual that we came up with, this is actually state law in now across 10 states. And intention of the state law is to kind of fill a gap in access to employee retirement plans, especially among smaller employers that have traditionally offered employer plans. So these are the 10 states that actually have this law in place now including my home state of New Jersey, so in New Jersey it's called a secure choice program and what this does is if you are an employer that does not offer a plan already, then you have to offer this secure choice program to all your employees. If you are a firm that has at least 25 workers and has been in business for at least two years. And once the workers get enrolled into this program they can opt out, but if they do not opt out, then the default contribution rate into this program is 3%, so 3% of your wages automatically contributed to this state plan. Okay so essentially it's making contributions to a retirement plan a default option and it's enrolling essentially all workers through this state law. So we can do this policy counterfactual in our data, which is suppose that in the last 10 years, this program was offered to the actual households in our data, so what that actually means is in each year in which these workers were working and received a W-2 suppose that we assume that they had access to an employer plan so we turn like these dummy variables that are zero if you had no access into ones, counterfactually assuming that they had access. Then we can compute the predicted change in retirement account participation, so among the zero to 20 income group 41% actually have access to retirement plans in the data, but under this policy counterfactual that would go up to 41% participate in retirement accounts currently, but under this policy counterfactual, that number would go up to 58%, so it's like an average treatment effect of 17 percentage points. So you can actually boost participation in retirement plans by 17 percentage points, if you actually make it mandatory that employers offering these plans. Now you'll notice that these treatment effects actually diminish as you consider higher income groups, so at the 80-100 percentile income group, the treatment effect is only one percentage point— why? Well these guys already participate regardless 98% of them already do. Moreover, they already have access, typically through their employer, so mandating that employers offer these plans has no effect because, first, they would participate anyway, and they already have access to their current employer. So kind of the key policy implications of this is the tax incentives that encourage retirement savings at the level of individuals might not work so well because we've been doing that forever and retirement plan participation is still low. But encouraging employers to offer these plans might have a bigger effect so, in particular, these mandates for employer staff retirement plans could have a potentially big effect on participation, especially for low income households.

Markus Brunnermeier: What is your most likely explanation for this finding, is it a purely behavioral story, inattention and that leads to this?

Motohiro Yogo: Yeah, so I think it's kind of a behavioral story like nudge is, if you tell people if you work for a small employer, you have to go to fidelity and open your own IRA, that's a pretty high cost, not only in terms of time, but also like kind of the barriers to doing, that is, you need to be kind of financially literate. But instead suppose the default option is regardless of which employer you work for in New Jersey, they're automatically going to take 3% of your payroll and contribute to the state retirement plan, then unless you've kind of fight against that default, then these guys are going to participate, so just that kind of story. Yeah it's very much important...

Markus Brunnermeier: It's still been the case that many people have been actively opt out, no?

Motohiro Yogo: That's exactly right, so this is not in our paper but there's a paper by Olivia Mitchell and co-authors that looked at the Oregon plan and they found a very high opt out rate, and that's actually consistent with like older papers by Bridget Madrian that you might be familiar with in the QJE, she showed that lots of workers opt out, but if, over time, if you allow these individuals to have access to retirement plans over decades they eventually on participate, so the take up rate is initially low and over a long period of time to kind of like gradually goes up so.

Markus Brunnermeier: You see, do see some firm effects as well that you know if my coworkers opt out, I will opt out too, we all opt out?

Motohiro Yogo: Yeah, that's a great question, I'll have to look back at the...

Markus Brunnermeier: You have the data for that, no?

Motohiro Yogo: I actually have the data for that, in terms of the W-2s, I can tell like yeah exactly that's a good idea. So from the W-2s, I can say I can observe what fraction of workers in your firm opt in, and we can see whether that explains the individual participation rates, yeah. So that would be a good idea.

Markus Brunnermeier: I would suspect that you know for my co-workers that they talk about it and work.

Motohiro Yogo: Right.

Markus Brunnermeier: Then opt out, so it might be bipolar almost.

59:22

Motohiro Yogo: It may also interact with firm size, maybe some like smaller and middle sized firms, then these pure effects might be more important. So those are things that we can definitely test for in our data. We haven't done that, but that's a great idea Marcus. Okay, actually I should mention kind of the— because we're on this topic, there are studies of retirement plans that are very influential but if you look at those studies they're about a single employer, typically the data select a single employer, and it doesn't allow you to look at whether these individuals had retirement accounts from previous jobs and so on, or whether they have IRAs and so on. So that's the advantage of tax data is you're able to really look at individuals

and households and look at their comprehensive financial planner not just add a single firm but the entire history of firms, as well as their individual retirement arrangement.

Markus Brunnermeier: Do you exploit the regional data here as well and does it matter whether I'm in a red or blue state or region?

Motohiro Yogo: We haven't done that so yeah we can try to test for group effects or peer effects so suppose I throw in as a specification, like the participation rate in the Princeton area. Does that explain individual participation in Princeton? That would be like a good way to test for you know pure effects, so we haven't done that at all, but it's definitely something that the data allows us to do. Very good, I don't know how much time we have left...

Markus Brunnermeier: We are running over time already.

Motohiro Yogo: Okay, so let me stop here.

Markus Brunnermeier: Perhaps you want to summarize, perhaps you can come back to the poll and say what surprised you the most by the participants or do you agree with what was predicted?

Motohiro Yogo: I'm very much encouraged by this particular audience, they indicated that they think it's a good idea for employers to offer a retirement plan. And I was very encouraged by those numbers, so, as you mentioned Marcus 2/3 either agree or strongly agree. Maybe, firstly, this is my fault that the wording in the question wasn't super clear as one someone pointed out in the chat, which is "what does it mean to require employers to offer?" That means that they don't employers don't necessarily have to contribute to the plan. All they're doing is giving an opportunity for workers to enroll in a plan, so it doesn't necessarily cost the employers actual money because they're not necessarily matching savings or doing a default contribution directly from the employer. What I mean by offer is that you just give them an opportunity to contribute from payroll with an option to opt out and that's what essentially the States are doing and it seems to be. The actual policy effects from such such plans– it's going to take a while for us to see the actual effects because again like just offering this in one year so I'm gonna boost savings or a participation immediately. These actually take decades for people to start contributing so the policies that came into effect are only in the last few years, but we'll have to wait another 10 years before we see the full effects, I think.

Markus Brunnermeier: I'm not sure why we have so much time to wait for another 10 years.

1:03:12

Motohiro Yogo: Yeah.

Markus Brunnermeier: But let me just– if we can say one more sentence, the fact that the bank participation went down for the less well off people over the last 10 years, what's the explanation for that? What's your favorite explanation for that? Then what policy measure would you do to correct for that phenomenon.

Motohiro Yogo: So that's where a lot of the questions that we get which are right on point and I think you also mentioned is maybe the reason it's declining is people are finding alternative

ways to make payments so like, are there PayPal accounts, Venmo accounts, are there maybe—Walmart apparently has some payment cards that allow you to like...

Markus Brunnermeier: Essentially then it doesn't show up in your text data, so people...

Motohiro Yogo: This is like substitution into more informal ways of making payments. So that's just one hypothesis, the other one is this is a post financial crisis thing so there are some papers on this topic that we cite in our people, which is post financial crisis, the regulatory environment for banks had changed in a way that was no longer profitable for banks to offer like free accounts and so on, so some of these fees went up in a way that disadvantaged low income. So one of the unintended consequences of the post crisis regulatory environment is that it might have made it more costly for low income households to maintain bank accounts. And if that's true, then that's kind of disturbing and then some people's papers argue this is true, they think that's what's going on, if that's true that's kind of a disturbing unintended consequence of the crisis and, if so, we need policy interventions.

Markus Brunnermeier: So some companies— don't know how in U.S., it is required for a bank to offer some basic bank account at the low fee, there's no such more in the United States is there?

Motohiro Yogo: No, that's another potential policy implication is to do a price control on fees for low account balances. Can I actually spend one minute just showing you one last thing that I wanted to end this touch upon your point about fintech, so can you see this? Yeah, so you can get to this from my website under research, household finance, financial inclusion, and then bank accounts/retirement accounts, we created these maps. So this is for bank accounts, and this is for retirement accounts, so what this map allows me to do is I can go to a zip code like 08540, Princeton New Jersey, which is actually among low income. Participation in retirement accounts is low in Princeton, and it's actually higher in West Windsor and Hopewell and we can keep zooming out, and we can see what's going on in New Jersey. For us, we have these very deep red areas in Trenton, near Elizabeth, in New York, and so on, so the reason I think this map to be important is this is like an opportunity for fintech firms to kind of try to target these regions that are red, because fintech could go in two directions, one is maybe they're just cream skimming off of the existing customers who would bank with traditional banks anyway. The alternative view is they're expanding the financial services to people who traditionally didn't have access to the services. So if it's the latter, then this type of map would be like very powerful in a sense that, okay, if I want to do a good thing, which is allow low income households to still open retirement accounts or bank accounts, then this map tells you exactly where those interventions are going to be more most effective. So I think this type of map is a powerful tool for both governments in fintech firms to boost participation and target those regions where participation is currently low.

Markus Brunnermeier: Very good, so it's always the case, like in Princeton, it's a high income area, but that's where the poor people are particularly badly banked so, again there's a high density of bank branches, I guess, in Princeton. And then help them, the poor people because, for what reason...

Motohiro Yogo: The idea behind fintech firms is like maybe traditional banks are too hard to use because of the fees and maybe the rigid structure of a bank, but maybe fintech firms are more nimble. All you need is a phone and very little and it doesn't take a lot to open an account. maintenance fees are lower, maybe they're able to reach the cell phones, maybe that's kind of

one view and if that's true, we're hoping this type of tool could help the fintech firms do what they're intending to do in a much more targeted way.

Markus Brunnermeier: Very good, I think that is opening a business model for a fintech company to do something good for society and hopefully somebody will pick it up, thanks a lot Moto. This was very fascinating and insightful and a first fascinating data set on top of it, but I guess with 360 million data points, and let me just do some little advertising in two weeks, we will have Alan Blinder presenting about soft and hard landings from the 1960s to the 2020s. So we will learn when the Fed was able to orchestrate a soft landing when it was not able to offer it to us, a soft landing. And it's based on a book he is writing, falling on Schwartz and Milton Friedman's book about the history of the Fed. He is essentially continuing that from the 1960s onwards and based on these books, he will present the two weeks about soft and hard landings, so I hope that all of you will come again in two weeks, it will be a Friday, instead of a Thursday. And thank you to all of you for participating and hope to see you soon again and particularly thank you to Moto for his fantastic presentation, bye bye.