Research Agenda: The Rich Interactions between Inequality and the Macroeconomy
Benjamin Moll, London School of Economics, November 2020

I originally wrote this piece for the Society of Economic Dynamics. It is also available online here, including a short teaser video summarizing the main topics covered below.

Introduction
One of the key developments in macroeconomics research over the last three decades has been the incorporation of explicit heterogeneity into models of the macroeconomy. As a result of taking micro data seriously, these theories study macroeconomic questions in terms of distributions of microeconomic variables like income or wealth rather than just aggregates. This approach is attractive for two reasons. First, empirically, it provides an integrated framework for making use of both micro and macro data. Second, conceptually, it provides a kind of “distributional macroeconomics” perspective, meaning an integrated perspective for analysing the distributional implications of macroeconomic trends, shocks or policies and the two-way interaction between distribution and the macroeconomy. My current research aims to contribute to this broader agenda.

The summary below describes some of my research projects in this area. These range from analysing the effects of policies like monetary policy or lockdown measures to tackle COVID-19 (Section 1) to studying potential drivers of rising income and wealth inequality like automation (Section 2) to the development of new methods for thinking about heterogeneity in macroeconomics (Section 3). I will conclude with what I view as some interesting open questions and avenues for future research (Section 4). All of the work I review here reflects collaborations – often over many years – with Yves Achdou, SeHyoun Ahn, Felipe Alves, Paco Buera, Andreas Fagereng, Xavier Gabaix, Jiequn Han, Martin Holm, Greg Kaplan, David Laibson, Jean-Michel Lasry, Pierre-Louis Lions, Peter Maxted, Gisle Natvik, Galo Nuño, Lukasz Rachel, Pascual Restrepo, Gianluca Violante, Tom Winberry, and Christian Wolf.

Writing this summary also prompted me to look back at some of the older RED research agenda pieces that can be found here. I highly recommend you do the same because they provide a very nice insight into the historical development of the agenda I described above, namely to develop an integrated approach to macroeconomic and distributional considerations as well as macro and micro data. See in particular the pieces by Victor Rios-Rull (2001), Tony Smith (2003), Kjetil Storesletten (2003), and Dirk Krueger and Fabrizio Perri (2005). Many of the themes in my research summary – as well as, more broadly, many of the recent developments in heterogeneous-agent macroeconomics – follow logically from the visions set out in these earlier pieces.

1. HANK Models for Macroeconomic Policy Analysis
Primarily in collaboration with Greg Kaplan and Gianluca Violante, one of the main themes of my work has been the development of richer and more empirically realistic models for macroeconomic policy analysis. These Heterogeneous Agent New Keynesian (HANK) models combine features from the heterogeneous agent (HA) literature, namely heterogeneity and incomplete markets, and the New Keynesian (NK) literature, namely nominal rigidities. They open
the door to studying distributional issues, business-cycle fluctuations, and stabilization policies, all within the same framework.

In “Monetary Policy According to HANK” (Kaplan, Moll and Violante, 2018) we revisit the transmission mechanism from monetary policy to household consumption in such a model, calibrated to yield empirically realistic distributions of wealth and marginal propensities to consume. In standard Representative Agent New Keynesian (RANK) models, monetary transmission is based almost entirely on intertemporal substitution. In contrast, in our HANK model this channel is small. Monetary policy nevertheless has sizeable real effects because of indirect effects, in particular those operating through a general equilibrium increase in labor demand and disposable incomes of high-MPC households. In Alves, Kaplan, Moll and Violante (2020) we follow up this work by examining in more detail the relative importance of different HANK model elements (e.g. unequal incidence of aggregate income fluctuations across households or the distribution of profits) for amplification or dampening of the response of aggregate consumption to a monetary shock.

Figure 1: Mechanisms through which monetary policy affects consumption in the HANK literature. For most up-to-date version, see http://benjaminmoll.com/HANK_summary/.

More broadly, there is now a flourishing literature with key contributions by many different researchers. Figure 1 provides an overview of the rich variety of transmission mechanisms through which monetary policy affects consumption in the HANK literature (the references are likely partly outdated, not meant to be comprehensive and I apologize for any omissions). Three broad lessons emerge relative to the RANK literature: (i) less important intertemporal substitution, (ii) more important direct income effects (e.g. through mortgage rates), and (iii) more important indirect general equilibrium effects. In short, a much richer picture of the monetary transmission mechanism than in RANK models with their emphasis on the Euler equation of a representative...
Income and wealth distribution have long played a central role in many fields of economics. The findings of the HANK literature show that they are also central in monetary economics: on the one hand, wealth distribution and portfolio composition determine the size and distribution of MPCs and these are important for the strength of the direct and indirect channels of monetary policy; on the other hand, how the distribution and composition of individual income comove with aggregate income determines the amplification of monetary policy and its redistributive consequences.

Apart from the lessons in the preceding paragraph the HANK literature offers few precise lessons about monetary transmission that are broadly generalizable. My view is that this is a good thing and simply reflects the complexity of the world we live in. Put differently, I would describe the overall message of the HANK literature as “the devil is in the detail, so nuance is required.” This is often in contrast to the answer to the same question one would obtain in a Representative Agent New Keynesian (RANK) model. For example, in a RANK model, for the central bank to boost aggregate consumption it is sufficient to influence current or expected future real rates: intertemporal substitution then ensures that expenditures will respond. In a HANK model, instead, the monetary authority must rely on equilibrium feedbacks that boost household income in order to influence aggregate consumption. Reliance on these indirect channels means that the overall effect of monetary policy may be more difficult to fine-tune by manipulating the nominal rate. The precise functioning of complex institutions, such as labor and financial markets, and the degree of coordination with the fiscal authority play an essential role in mediating the way that monetary interventions affect the macroeconomy. This type of finding appears frequently in the HANK literature. To give a concrete example, Alves, Kaplan, Moll and Violante (2020) and Auclert, Rognlie and Straub (2020) find that investment plays a key role in the transmission of monetary policy to consumption, precisely because it triggers indirect effects working through household labor income. This is an example of the type of intuitive effect that is simply absent from a model organized around the Euler equation of a representative agent.

In “Present Bias Amplifies the Household Balance-Sheet Channels of Macroeconomic Policy” (2020) with David Laibson and Peter Maxted, we venture further into trying to understand the implications of the complex financial planning problems faced by real-world households, and of the psychological factors that influence them. To this end, we study the effect of monetary and fiscal policy in a heterogeneous-agent model where households have present-biased time preferences, a form of dynamic inconsistency that has received empirical support in both laboratory and field studies. The model features a liquid asset and illiquid home equity, which households can use as collateral for borrowing. Because present bias substantially increases households’ marginal propensity to consume (MPC), present bias increases the impact of fiscal policy. Present bias also amplifies the effect of monetary policy but, at the same time, slows down the speed of monetary transmission. Interest rate cuts incentivize households to conduct cash-out refinances, which become targeted liquidity-injections to high-MPC households. But present bias with naive beliefs of the type we model also introduces a motivation for households to procrastinate on refinancing their mortgage. Intuitively, naive present bias implies that households will delay completing immediate-cost delayed-reward tasks such as mortgage refinancing, which involves lots of up-front paperwork. Naive households will continually delay refinancing, all the while (counterfactually) believing that the task will get done in the near future. As a caveat and to
tie the analysis back to Figure 1, I should be clear that our model is thus far set in partial equilibrium. As such, it can speak to only the direct effects of both monetary and fiscal policy. A natural next step is to embed our present-biased households in a full general equilibrium, i.e. a full-blow HANK model, so as to speak to indirect effects like those in Figure 1.

Heterogeneous agent models can, of course, also be used to study other policies besides regular monetary and fiscal policy. At the time of writing, we are in the midst of the COVID-19 pandemic, and some of the most important policy questions of the day are concerned with the appropriate policy responses to it and in particular the question to what extent different policies involve a trade-off between saving lives and preserving livelihoods. In “The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S.” (Kaplan, Moll and Violante, 2020), we contribute to this debate by quantifying this trade-off, focusing on the distributional effects of the pandemic and associated policy responses, across different types of workers and households. One of our main arguments is that the choice governments face when designing policy is not just between lives and livelihoods, as is often emphasized, but also over who should bear the burden of the economic costs.

To make this argument, we integrate an expanded SIR model of virus spread into a macro model with realistic income and wealth inequality, as well as occupational and sectoral heterogeneity. Our starting point is that many of the individuals who are most financially exposed to the pandemic are also the most financially vulnerable. A key determinant of economic exposure is occupation. Socially facing workers who cannot work remotely (such as waiters and hairdressers), have experienced especially large drops in earnings. In contrast, the earnings of workers in occupations that produce goods and services that do not require social interaction and have high flexibility to work from home (such as lawyers, academics, and finance professionals) have been left relatively unscathed. In our model, as in the data, the most exposed occupations also have the lowest liquid wealth to buffer such income shortfalls. Our model therefore predicts that the welfare losses due to the pandemic have been extremely unequal across the population. By the same token, there is scope for economic and health policies, with appropriate patterns of redistribution, to both contain the virus and mitigate its economic effects. We summarize our findings through a “distributional pandemic possibility frontier” (PPF), which shows the distribution of economic welfare costs associated with the different aggregate mortality rates arising under alternative containment and fiscal strategies.

We use our model to evaluate the CARES act, the large fiscal policy package implemented in the U.S. in the spring of 2020 and find that it was quite effective at alleviating economic hardship, in particular mitigating economic welfare losses by around 20% on average. Our integrated framework can also be used to find alternative policies that alleviate the tradeoff between lives and livelihoods. For example, we find that Pigouvian taxes on social consumption and work in the workplace, with revenues rebated to the workers employed in rigid and social-intensive occupations, is very effective at attaining a more favorable average tradeoff. However, they come at the cost of more unequal economic welfare losses, a feature which would have to be appropriately managed through fiscal redistribution for these policies to be feasibly implemented in practice.
As I hope these examples demonstrate, HANK models have the potential to be useful for macroeconomic policy analysis and they already seem to factor into policy decisions, e.g. the shift to the Fed’s new monetary policy framework (Powell, 2020; Feiveson et al., 2020). Before they can be used on a day-to-day basis, economists will however need to accumulate more experience in solving, calibrating, estimating – or plainly “using” – this class of models. Partially to try to help in this regard, in 2019 Greg Kaplan and I started teaching a four-day “Master Class” for central bankers and economists from other policy institutions. Our goal is to help familiarize economists with the state-of-the-art tools for incorporating income and wealth distributions into macroeconomic models, and the main policy lessons that have emerged from these models. We unfortunately had to cancel the 2020 class due to the COVID-19 pandemic but we hope to resume teaching it in 2021.

2. Theories of Rising Income and Wealth Inequality

Over the past forty years, economic growth in the United States has been unevenly distributed: income percentiles corresponding to the lower half of the distribution have stagnated while those at the top have sharply increased (e.g. Census Bureau, 2015, and Piketty, Saez and Zucman, 2018). Similar trends have occurred in many other advanced economies. A second theme of my work has been trying to understand these trends.

Since Pareto (1896), it has been well known that the upper tail of the income distribution follows a power law, or equivalently, that top inequality is “fractal,” and the rise in top income inequality has coincided with a “fattening” of the right tail of the income distribution. That is, the “super rich” have pulled ahead relative to the rich. This rise in top inequality requires an understanding of the forces that have led to a fatter Pareto tail. In “The Dynamics of Inequality” (Gabaix, Lasry, Lions and Moll, 2016) we show that the most widely used theories of the observed fat tails of these distributions, which build on a random growth mechanism, generate transition dynamics that are too slow relative to those observed in the data. We then suggest two parsimonious deviations from the canonical model that can explain such changes: “scale dependence” that may arise from changes in skill prices, and “type dependence,” that is, the presence of some “high-growth types.” While our work stops short of assessing concrete economic mechanisms put forth in the public debate – is the rise in top inequality due to: technical change, superstars, rent-seeking, globalization, and so on? – it provides some structure for economists trying to develop theories of fast changes in inequality. Economic mechanisms that can generate type- or scale-dependence (or both) are particularly promising, for example theories involving “superstar” phenomena.

One potential driver of rising income inequality that is often cited by pundits and policy makers alike is technical change, and in particular the automation of tasks performed by labor, and a large literature in macro and labor economics has studied how technology and automation affect the distribution of labor incomes. But not all income is labor income and capital income is an important income source, particularly at the top of the distribution where incomes have increased the most. Existing theories therefore paint an incomplete picture of technology’s implications for overall income inequality. In “Uneven Growth: Automation’s Impact on Income and Wealth Inequality” (Moll, Rachel, Restrepo, 2019) we therefore develop a theory that links technology to the personal income and wealth distributions – and not just that of wages – and use it to study the distributional effects of automation.
Our main argument is that technology affects not only relative wages but also asset returns and this can have substantial distributional effects. This argument has two parts. First, automation directly contributes to income inequality by increasing returns to wealth and the concentration of capital ownership. Second, relative to theories in which returns are unaffected, automation is also more likely to lead to stagnant wages and therefore stagnant incomes at the bottom of the income distribution (even in the long run). The key for understanding both parts of the argument is that long-run capital supply in our model is upward-sloping. Automation increases the demand for capital relative to labor and because supply is upward-sloping, this demand shift permanently increases returns to wealth. Importantly, this is in contrast to many workhorse theories of capital accumulation, in particular variants of the neoclassical growth model, in which long-run capital supply is perfectly elastic and therefore returns don’t budge in response to demand shifts. Paraphrasing this logic: in workhorse theories, if “robots” increasingly outperform labor, in the long-run this always benefits workers; in contrast in our theory, this benefits people owning lots of robots. This paper is still work in progress and so I will here stop short of reporting quantitative model predictions and comparing them to data – these will be different in the paper’s next iteration. For example, the next iteration will feature an extension of our model with multiple assets and returns which allows us to speak to empirical trends on asset returns, in particular that treasury rates have declined over time while the return to other assets, including US business capital and equity have increased (e.g. Caballero, Farhi and Gourinchas, 2017; Gomme, Ravikumar and Rupert, 2011).

3. Continuous-Time Methods for Macro Models with Distributions

A unifying thread in all of my research is that I try to develop better methods for thinking about heterogeneity in macroeconomics. This theme goes back to Moll (2014) and Buera and Moll (2015). “Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach” (Achdou et al., 2020) shows that, when recast in continuous time, incomplete-market models can be conveniently solved as systems of partial differential equations. This approach allows for both a tighter theoretical characterization and more efficient computations than traditional discrete-time methods. The model with two assets and kinked adjustment costs developed in Kaplan, Moll and Violante (2018) and the model of mortgage refinancing with present bias in Laibson, Maxted and Moll (2020) provide illustrations of the usefulness of these methods. Also see Bornstein (2020), McKay and Wieland (2020) and Guerrieri, Lorenzoni, and Prato (2020).

In “When Inequality Matters for Macro and Macro Matters for Inequality” (Ahn et al., 2017) we further extend this methodology to handle aggregate uncertainty and develop a computational toolbox for numerically solving such models. If you are looking for a computational method to solve your heterogeneous agent model with aggregate risk, I recommend that you additionally take a look at the promising technical contributions by Fernandez-Villaverde, Hurtado and Nuño (2020) and McKay and Wieland (2019) (the latter building on Auclert, Bardóczy, Rognlie and Straub’s (2019) “sequence-space approach”). Finally, Nuño and Moll (2018), Nuño and Thomas (2019) and González, Nuño and Thaler (2020) extend the method to compute social optima which is a challenging problem because the planner’s state variable is typically a distribution.
4. Open questions and avenues for future research

Let me conclude with what I view as three interesting avenues for future research. My hope is that this may be useful for young researchers who would like to work in the field. I chose these three avenues because I think they are particularly under-researched; there are of course many other important avenues related to the themes I discussed above, for example better understanding the linkages between heterogeneity and aggregate demand (Yellen, 2016) or optimal monetary and fiscal policy with heterogeneity, but these have already received relatively more attention. Finally, I should also be clear that what follows is necessarily much more speculative than the rest of this piece, and that some of these avenues may well turn out to be dead ends.

4.1 Heterogeneous-agent macro as a gateway to “behavioral macro”

A key part of the philosophy of heterogeneous-agent macro is to try to build things “from the ground up” and to take seriously empirical evidence on individual behavior. The household finance and behavioral economics literatures have documented a number of empirical patterns that are hard to rationalize with models of fully optimizing behavior, e.g. with regard to pension saving or mortgage refinancing, and have suggested “behavioral” deviations to make sense of these. A logical next step is therefore to incorporate the insights from these literatures into our macro models and to ask whether they change the models’ predictions for things like the transmission of monetary and fiscal policy, whether they can help generate more realistic business-cycle fluctuations, and so on. In summary, heterogeneous agent macro is a natural gateway to “behavioral macro.”

The paper with David Laibson and Peter Maxted on the implications of present bias for monetary and fiscal policy I discussed in Section 1 takes a step in this direction. Also see Maxted (2020) and related work on the implications of temptation preferences by Attanasio, Kovacs and Moran (2020). There are many other interesting deviations from rationality or rational expectations that it may be worthwhile to incorporate into heterogeneous-agent macro models so as to gauge their implications. These include rational inattention (Woodford, 2002; Sims, 2003), sticky information (Gabaix and Laibson, 2002; Mankiw and Reis, 2002, 2007; Auclert, Rognlie, and Straub, 2020), “sparsity” (Gabaix, 2019), learning from experience (Malmendier and Nagel, 2011, 2016; and see Lei, 2020, for a nice application), and “near-rationality” (Kueng, 2018; Lian, 2020). An important guiding principle in such an endeavor should be to limit attention to those deviations from rationality or rational expectations for which there is solid empirical support from laboratory or field studies and, going forward, to think even harder about different deviations’ testable implications. Such an evidence-driven approach will hopefully prevent “behavioral heterogeneous-agent macro” from turning into a “zoo” with too many competing theories to keep track of.

4.2 Asset price changes and wealth inequality

Empirically a large chunk of increasing wealth inequality is due to changing asset prices (see e.g. Kuhn, Schularick and Steins, 2020, and Martínez-Toledano, 2020). This raises a number of interesting questions.

One of these is: If a large fraction of the increase in wealth inequality is due to asset price changes, should we care? Do those whose wealth increases due to rising asset prices also benefit in welfare terms? Or are such capital gains just “paper gains”? In a nutshell, do asset price changes that increase wealth inequality also increase welfare inequality? In a recent NBER Macro Annual
comment on a very nice related paper by Hubmer, Krusell and Smith (Moll, 2020) I tried to think through this questions using a simple two-period toy model. The answer coming out of this simple exercise was “it depends”, with two more concrete lessons. The first lesson is that the source of capital gains matters, in particular whether rising asset prices are primarily due to rising dividends or falling discount rates. The second lesson is: “investment plans” matter, i.e. whether investors intend to buy, sell, or keep their portfolios unchanged. To take an extreme example, if the only reason asset prices increase is falling interest rates and investors just live off dividends, nothing happens to their welfare. But in many other cases also welfare inequality increases. This discussion immediately suggests two open questions. First, to what extent do these findings carry over to more realistic model environments? Second, the empirical question: which of these cases is most relevant?

The more general point is that asset price changes are not merely pesky “valuation effects” to be treated as residuals but that they are empirically important for understanding wealth inequality dynamics and also raise some interesting conceptual issues. It is therefore good to see that an emerging theoretical literature is starting to emphasize portfolio choice and asset price changes as drivers of wealth inequality (e.g. Garleanu and Panageas (2020), Hubmer, Krusell and Smith, 2020; Gomez, 2018; Gomez and Gouin-Bonenfant, 2020). Similarly we need more empirical evidence, e.g. on how net asset purchases and capital gains differ across the wealth distribution (Bach, Calvet and Sodini, 2018; Fagereng, Holm, Moll and Natvik, 2019) and this evidence needs to be tied in with the new models.

4.3 Heterogeneity, Bubbles and Crashes
More than ten years after the Great Recession, macroeconomists still do not have a sound understanding of the root causes of infrequent but large economic crises. In his classic treatment of historical financial crises, Kindleberger (1978) writes: “By no means does every upswing in business excess lead inevitably to mania and panic but the pattern occurs sufficiently frequently and with sufficient uniformity [...] In the manic phase, people of wealth or credit switch out of money to borrow to buy real or illiquid financial assets. In panic, the reverse movement takes place, for real or financial assets to money or repayment of debt, with a crash in the prices of commodities, houses, buildings, land, stocks, bonds -- in short, whatever has been the subject of the mania.” An open question is whether one can write down a version of this story that lives up to the standards of modern heterogeneous-agent macro: quantitative theory disciplined with the relevant micro data.

An interesting avenue for future research is therefore to try to develop quantitative macro theories of bubbles and crashes that are consistent with empirical household balance sheets and portfolio rebalancing in booms and busts. One reason this will be challenging is computational constraints. Capturing Kindleberger’s narrative would likely require a heterogeneous-agent model with: (i) time-varying asset prices, (ii) multiple assets, (iii) non-linearities at the micro level, (iv) non-linearities at the aggregate level so as to deliver infrequent but large economic crises. Computing such a model is hard. Even more important are conceptual questions, e.g. how flows in and out of asset markets can have large impacts on prices. Tools like those of Fernandez-Villaverde, Hurtado and Nuño (2020) and ideas like those of Gabaix and Koijen (2020) may be promising in this regard.
References


