

Economics for Inclusive Prosperity: An Introduction

Suresh Naidu, Dani Rodrik, and Gabriel Zucman¹

Background

We live in an age of astonishing inequality. Income and wealth disparities between the rich and the poor in the United States have risen to heights not seen since the gilded age in the early part of the 20th century, and are among the highest in the developed world. Median wages for American workers remain at 1970s levels. Fewer and fewer among newer generations can expect to do better than their parents. Organizational and technological changes and globalization have fueled great wealth accumulation among those able to take advantage of them, but have left large segments of the population behind. U.S. life expectancy has declined for the third year in a row in 2017, and the allocation of healthcare looks both inefficient and unfair. Advances in automation and digitization threaten even greater labor market disruptions in the years ahead. Climate change fueled disasters increasingly disrupt everyday life. Greater prosperity and inclusion both seem attainable, yet the joint target recedes ever further.

This is a time when we need new ideas for policy. We think economists, among other social scientists, have a responsibility to be part of the solution, and that mainstream economics – the kind of economics that is practiced in the leading academic centers of the country – is indispensable for generating useful policy ideas. Much of this work is already being done. In our daily grind as professional economists, we see a lot of policy ideas being discussed in seminar rooms, policy forums, and social media. There is considerable ferment in economics that is often not visible to outsiders. At the same time, the sociology of the profession – career

incentives, norms, socialization patterns – often mitigates against adequate engagement with the world of policy, especially on the part of younger academic economists.

We believe the tools of mainstream economists not only lend themselves to, but are critical to the development of a policy framework for what we call “inclusive prosperity.” While prosperity is the traditional concern of economists, the “inclusive” modifier demands both that we consider the whole distribution of outcomes, not simply the average, and that we consider prosperity broadly, including non-pecuniary sources of well-being, from health to climate change to political rights. The policy briefs that accompany this overview offer a range of policy recommendations, drawn from labor economics, public economics, international economics, financial economics, etc.² Importantly, we hope this collective effort amounts to more than a discussion of specific policy prescriptions in different domains of economics. Our claim is that there are overarching themes and commonalities that taken together provide a coherent overall vision for economic policy that stands as a genuine alternative to the market fundamentalism that is too often (and in our view, wrongly) associated with mainstream economics. We strive for a whole that is greater than the sum of the parts.

We shall discuss these broader themes and the connecting narratives that emerges later in this essay. We begin by discussing in greater detail the motivation behind the project and the role that we see economics in crafting an alternative to the status quo.

Why we are doing this

The idea for this initiative developed following a workshop that the three of us attended during the first half of 2018. It was one of those multi-disciplinary meetings that have become increasingly common recently, on “new thinking beyond neoliberalism” and similar themes. The organizers had brought together historians, political scientists, sociologists, and legal scholars alongside economists. As is usual in such meetings, participants agreed that the prevailing policy framework had failed society, resulting in monumental and growing gaps in income and wealth. All of us were horrified by the illiberal, nativist turn our politics had taken, fueled in part by these chasms. There was consensus that we needed to develop a genuine alternative – a set of policies that were both effective and inclusive, responding to legitimate grievances without sowing deeper societal divisions.

Any economist who sits in such a meeting will eventually find himself or herself on the defensive. For in the eyes of many, the turn towards neoliberalism is closely associated with economic ideas.³ Leading economists such as Friedrich Hayek and Milton Friedman were among the founders of the Mont Pelerin Society, the influential group of intellectuals whose advocacy of markets and hostility to government intervention proved highly effective in reshaping the policy landscape after 1980. Deregulation, financialization, dismantling of the welfare state, de-institutionalization of labor markets, reduction in corporate and progressive taxation, and the pursuit of hyper-globalization – the culprits behind rising inequalities – all seem to be rooted in conventional economic doctrines. The discipline’s focus on markets and incentives, methodological individualism, mathematical formalism, and passion for causal identification all seem to point towards the status quo and stand in the way of meaningful economic and social reform. In short, neoliberalism appears to be just another name for mainstream economics.

Consequently many non-economists view the discipline of economics if not with outright hostility, at least as part of the problem. They believe the teaching and practice of economics has to be fundamentally reformed for the discipline to become a constructive force.

And there are, indeed, legitimate reasons for the discontent with economics, the way it is too often practiced and taught. Conservative foundations and

think tanks have monopolized the banner of economics in policy circles, pushing the view that there is a steep efficiency-equality trade-off, and normatively prioritizing economic growth. Students who do not pursue further training leave undergraduate courses thinking that economics means that “markets always work”. Conservatives tend to deploy “economics” as a justification for preferred policies, while liberals are seen to be insensitive to the requirements for prosperity.

But our own take goes beyond this common view and is substantially different from it. Many of the dominant policy ideas of the last few decades are supported neither by sound economics nor by good evidence. Neoliberalism – or market fundamentalism, market fetishism, etc. — is a perversion of mainstream economics, rather than an application thereof. And contemporary economics research is rife with new ideas for creating a more inclusive society. But it is up to us economists to convince their audience about the merits of these claims. That is why we have embarked on this project. The initial set of policy briefs that accompany this introduction is our first step. We hope they will stimulate and accelerate academic economists’ sustained engagement with creative ideas for inclusive prosperity.⁴

Economics is an ally of inclusive prosperity

How do we square non-economists’ perception with our claim that economics is part of the solution?

Economists study markets (among other things) and they naturally feel a certain pride in explaining the way they operate to those who lack their specialized knowledge. When markets work well, they do a good job of aggregating information and allocating scarce resources. The principle of comparative advantage, which lies behind the case for free trade, is one of the profession’s crown jewels – both because it explains important aspects of the international economy and because it is, on the face of it, so counter-intuitive. Similarly, economists believe in the power of incentives, because they have evidence people respond to incentives and they have seen too many well-meaning programs fail on account of not having paid adequate attention to the creative ways in which people behave to realize their own goals.

At the same time, contemporary economics is hardly a paean to markets and selfishness. The typical course in microeconomics spends more time on market failures and how to fix them than on the magic of competitive markets. The typical macroeconomics course focuses on how governments can solve problems of unemployment, inflation, and instability rather than on the “classical” model where the economy is self-adjusting. The typical finance course revolves around financial crises, excessive risk-taking, and other malfunctions of financial systems. In fact, the standard competitive equilibrium model in which free markets are maximally efficient (even if still not necessarily socially optimal, in view of distributional concerns) is the dominant framework only in introductory economics courses. Serious students of economics quickly move away from it.

Economics remains somewhat insular within social sciences because of its methodological predilections: methodological individualism, model-based abstraction, mathematical and statistical formalism. But in recent decades economists have reached out to other disciplines and have incorporated many of their insights. Economic history is experiencing a revival, behavioral economics has put *homo economicus* on the defensive, and the study of culture has become mainstream. Distributional considerations are making a comeback at the center of the discipline. Economists have been at the forefront of studying the growing concentration of wealth, the costs of climate change, concentration of important markets, the stagnation of income for the working class, and the changing patterns in social mobility.

Economists often have a bias towards market-based policy solutions, sustained by a demand for identifying precise market failures as a precondition for policy interventions. But the science of economics has never produced pre-determined policy conclusions. In fact, all predictions and conclusions in economics are contingent: if these and these conditions hold, then these outcomes follow. The answer to almost any question in economics is “it depends,” followed by an exegesis on what it depends on and why.⁵ Back in 1975, in a collected volume titled *International Trade and Finance: Frontiers for Research* an economist wrote: “by now any bright graduate student, by choosing his assumptions ... carefully, can produce a consistent model yielding just about any policy recommendation he favored at the start” (Diaz Alejandro, 1975). Economics has become even richer in the intervening four decades. We might say, only slightly facetiously, that today the graduate

student need not even be that bright!

Moreover, economics research has become significantly more applied and empirical since the 1990s. The share of academic publications that use data and carry out empirical analysis has increased substantially in all sub-fields within economics, and currently exceeds 60 percent in labor economics, development economics, international economics, public finance, and macroeconomics (Angrist et al., 2017). This is important because systematic empirical evidence is a disciplining device against ideological policy prescriptions embedded in preconceived theorizing. The empirical bent of economics makes it more difficult to ignore inconvenient facts, when real world markets do not behave like textbook ones. It is harder to idolize markets when research finds international trade produces large adverse effects on some local communities, minimum wages do not reduce employment, or financial liberalization produces crises rather than faster economic growth – just to point out a few empirical findings from the recent economic literature.

Economics does have its universals, a set of higher-order principles associated with efficiency and generally presumed to be conducive to superior economic performance: market-based incentives, clear property rights, contract enforcement, macroeconomic stability, prudential regulation, and so on. But these principles are compatible with an almost infinite variety of institutional arrangements. Each of these institutional arrangements – rules of the game — produces a different distributional outcome. And how it contributes to overall prosperity depends on the suitability to the specific context at hand. This is a recipe for comparative institutional analysis of economic performance, and no glib “markets work” slogans follow from it. The abstraction with which economists perceive complex bundles of institutions also gives practitioners tools to help design large scale alternatives; from precision tweaks to the tax code to full-blown visions of post-capitalist societies.

Consider the simplest economic setting of a perfectly competitive market economy. When an economist draws a supply-and-demand diagram on the black board, she may not list all the institutional prerequisites that lie behind the two curves. Firms have property rights over their assets and can enforce their contracts with suppliers. They have access to credit, can rely on public infrastructure such as transportation and power, and

are protected from thieves and bandits. Their employees accept the terms of employment and show up at work each day. Consumers have all the information they need to make reasonable choices. They are reasonably confident that firms do not cheat them. There is a stable unit of value and means of exchange for buying and selling goods.

Clearly markets rely on a wide range of institutions; they are “embedded” in institutions, as Karl Polanyi would say. But how should those institutions be designed? Take property rights, for example. The Coase theorem suggests it does not matter for efficiency how property rights are allocated as long as transaction costs are zero. But the caveat does a lot of work here, even if we focus only on efficiency: clearly transactions costs matter greatly. So we must make choices. Should a job belong to a company, a worker, or a combination? Perhaps the company itself should be owned by a third party -- a local government entity, say -- and simply ensure incentive compatibility for managers and workers. You might think this is crazy, but China has eked unprecedented rates of economic growth out of such a property-rights regime. We can think of many other variants. Perhaps employers should have property rights (for a fixed period) only over new assets they create, with existing assets distributed among other claimants. That too sounds crazy, unless we realize that is exactly what the patent system does, giving innovators temporary ownership over new “intellectual property.” Perhaps government should retain part ownership of new technologies, on behalf of the general public, since so much of innovation relies on public infrastructure (public R&D and subsidies, higher education, the legal regime, etc.). Distributional concerns add to the choices that need to be made. Which among these (and other) possibilities we should favor depends both on our ultimate objectives and the potential fit with local context.

As we grapple with new realities created by digitization, demographics, and their impacts on labor markets, such questions about the allocation of property rights among different claimants become crucial. Economics does not necessarily have definite answers here. Nor does it provide the appropriate distributional weights (how do we weight the returns to workers, employers, and the government, and what procedural and deontological constraints should be respected). But it supplies the tools needed to lay out the tradeoffs, thus contributing to a more informed democratic debate.

The same kind of institutional indeterminacy pervades all other policy domains. Which labor market institutions minimize job insecurity without jeopardizing employment creation? How do we best provide social protection without blunting economic incentives? What kind of financial regulations ensure financial stability without blocking financial innovation? What kind of monetary and fiscal rules are best for an open economy? Once again, economics does not provide a fixed answer to these questions. Instead, it highlights the potential consequences of different arrangements.

Economists have a powerful theoretical machinery that allows them to think in abstract terms about such matters. So they are well positioned to develop innovative institutional arrangements that go beyond the already considerable variety that exists in the world today. Welfare or labor-markets arrangements, say, differ greatly across the developed world. There is much that the U.S. can learn from experiments elsewhere. But plausible institutional diversity is not limited to existing practices. We can – and will need to – to develop new institutions. Nothing in free markets guarantees that growth will be equitable or globalization sustainable. There’s always a need to design policies and institutions to make inclusive prosperity possible and globalization sustainable politically and economically. Economists’ imagination is crucial to the task.

Economists’ habits are to blame too

The misunderstanding of what economics is (and what economists do) is compounded by the way economists frequently engage in public debates. Too many economists believe their quantitative tools and theoretical lenses are the only ones that count as “scientific,” leading them to dismiss disciplines that rely more on qualitative analysis and verbal theorizing. Many economists feel they need to take the side of markets, because no-one else will do so and because doing otherwise might “provide ammunition to barbarians.” And even when they recognize market failures, they worry government action will make things worse. As a result, many of the discipline’s caveats are swept under the rug. And economists get labeled as cheerleaders for free markets and hyper-globalization.

There is often a naïve political economy at play here,

with the implicit assumption that self-interested pressure groups and rent-seekers – the so-called barbarians -- are represented only on one side of a policy question.⁶ In reality, every market equilibrium, with or without public action, creates winners and losers. These groups necessarily try to bend outcomes to their liking. Neoliberalism certainly has had its own powerful lobbies. Free-market oriented policies since the 1980s have been hijacked by their own special interests, as we can see in corporate taxation or trade agreements for example. Good policy cannot be abstracted from politics and has to be designed by taking its likely effects into account. This is as true for policies that purportedly try to take the government out of the market as it is for policies that broaden the government's role.

Economists often get too enamored of first-best benchmarks within a model tailored to study a narrow set of issues. This leads them to focus on the direct efficiency consequences in the area under focus, at the expense of potential complications and adverse implications elsewhere. A growth economist will analyze policies that enhance technology and innovation without worrying about labor market consequences. A trade economist will recommend reducing tariffs, and assume that devising compensatory mechanisms for the losers is somebody else's job. A finance economist will design regulations to make banks safe, without considering how these may interact with macroeconomic cycles. Many policy failures – the excesses of deregulation, hyper-globalization, tax cuts, fiscal austerity – can be traced to such first-best reasoning. To be useful, economists have to evaluate policies in the totality of the context in which they will be implemented, and consider the robustness of policies to many possible institutional configurations and political contingencies. As Avinash Dixit (2009) puts it, “the world is second-best at best.”

Some common themes in the policy essays

All of the participants in this project are academic economists, working in broadly mainstream subfields. Some have worked in government; most have not. Some have engaged in writing broadly for a non-academic audience; most have not. They are researchers who believe sound scholarship is indispensable to show the way to inclusive prosperity. They are all economists of the real world, who understand that we live in a

second-best world rife with market imperfections, and in which power matters enormously in shaping market outcomes.

In such a world the competitive model is rarely the right benchmark for understanding the problems and suggesting solutions. We must instead search for alternative models. This requires an empirical orientation, an experimental mind set, and a good dose of humility – to recognize the limits of our knowledge.

The policy proposals in these essays reflect economic reasoning and contemporary evidence on a variety of market failures, from international trade to insurance to capital and labor markets. Shot through the proposals is the sense that economies are operating well inside the justice-efficiency frontier, and that there are numerous policy “free-lunches” that could push us towards an economy that accords with our moral intuitions without sacrificing (and indeed possibly enhancing) prosperity. Taking contemporary economics seriously is consistent with recommending fairly dramatic structural changes in American economic life.

Many of the proposals involve efficiency-and-equality enhancing interventions in markets well known to be rife with market failure, such as labor markets (Dube and Naidu), credit markets (Admati and Mian), insurance markets (Black and Rothstein), and markets for innovation (Korinek). While the theoretical basis for market failures in these domains has been apparent for some time, the empirical importance of the various failures has been made only recently.

For example, while the minimum wage debate continues, there is a consensus that it is not an effective tool for intervening in labor markets with wages higher than say, the 30th percentile. Other labor market institutions are needed to take advantage of free lunches created by monopsony and other labor market failures in the segment of the labor market where most workers find themselves. Dube proposes a system of wage boards, similar to the Australian system, where either administrators or tripartite boards negotiate wages at the industry-occupation-region level, thus setting minimum wages throughout the distribution. He finds that wage inequality would significantly fall as a result. Naidu discusses the more traditional American labor movement, and possibilities for economics to help organized labor overcome some of the limitations of the current U.S. industrial relations institutions.

In the domain of capital markets, both Admati and Mian stress the systemic risk produced by the current system. Mian discusses the role that inequality, together with capital flows from oil-rich countries and Asia, has played in generating a “glut” of savings, pushing down the real interest rate and increasing systemic risk. Admati looks at the banking sector, showing how banks, uniquely among financial institutions, are overexposed to debt, making them more vulnerable to bankruptcy and again, a threat to stability. Both authors point to a variety of macroprudential regulatory options, with Mian emphasizing credit contract repayments that are contingent on the aggregate state of the economy, and Admati favoring capital requirements and tax reforms that make debt look less attractive.

Two of the proposals speak directly to how the size for government can be increased in a sustainable and prosperity-enhancing way. Zucman’s proposal shows an ingeniously simple path out of international tax competition, where countries no longer have to bid for multinational investment by slashing corporate taxes. Zucman proposes taxing multinationals by allocating their global profits proportionally to where they make their sales. While companies can easily relocate profits or production to low-tax jurisdictions today, sales are much harder to manipulate. His reform would thus make it possible to tax the very winners of globalization—probably a necessary condition for globalization to be sustainable in the long run.

Black and Rothstein provide a contemporary restatement of an old idea: government should provide public goods and social insurance, and root this argument in the best modern economics. For example, education requires government provision because parents cannot borrow against the earnings of their children (and children happen generally before the peak income of the parents). The benefits of education are also in the far future, and are associated with externalities in crime, citizenship, and innovation. All this militates in favor of government provision of education. Social insurance mitigates the widespread and well-known failures in insurance markets, in the form of unemployment insurance, social security, and health insurance.

Korinek takes up the increasingly important question of how new technologies affect labor markets and the distribution of income. The direction of technological change is not exogenous, he argues, and it depends on the incentives set both by markets and by governments.

In particular, innovators may over-estimate the social cost of labor, investing too much in technologies that replace labor. Governments routinely intervene in the process of innovation, for example to encourage green technologies. Korinek proposes that they similarly steer technology in the direction of innovations that have desirable distributive properties. Promoting AI systems that complement and augment the cognitive abilities of workers – along with mechanisms that ensure workers retain substantial part of the surplus generated – would be an example. Korinek also discusses how inelastic, complementary factors such as land or specialized skills might be taxed in response to technological change, and how the value of monopolies granted by the patent system is intrinsically inequalitarian, as it transfers income from consumers to owners of firms.

Another way to look at a slice of the proposals is via Karl Polanyi: to work well, crucial markets (e.g. the “fictitious commodities” of labor, land, and capital) must be embedded in non-market institutions, the “rules of the game” supplied by government. Rodrik, for example, shows that trade agreements ought to include clauses that prevent competition on “unjust” margins; and Dube shows that wage boards setting market-specific minima could compress wages a lot, with much more refined targeting than a blunt, economy-wide minimum wage. Mian shows how inequality generates instability in financial markets, but also how private macro-prudential contracting is thwarted both because there is an aggregate externality as well as specific tax and regulatory structures (e.g. Basel III risk weighting). Rodrik’s proposal is distinctive in that it gives an explicitly pro-social justification for restrictions on trade, not trying to clothe the protectionism in terms of ameliorating some other externality or market failure. Rodrik’s “social safeguards” would give countries a claim, justified by broad social support, on trade authorities that a restriction on trade is necessary to maintain the domestic social contract. This proposal is indicative of the commitments of many of the members of EFIP; a willingness to subordinate textbook economic efficiency to other values such as democratic rule and egalitarian relationships among citizens.

Finally, some of the proposals propose fixing non-market institutions with ideas from economics. Importantly for any policy proposals in 2018, democratic political economy must be considered, where people’s influence on policy is roughly equal and political preferences are arrived at through open, well-informed public debate.

Too many policy ideas break on the rock of government capture by special interests or systematically distorted presentations in the media. Ethan Kaplan's proposal draws on a few decades of empirical political economy to suggest policies that could drastically alter the balance of political influence in the United States. Suresh Naidu's proposal hints at ways mechanism design and behavioral economics can be mobilized to ease the pervasive collective action problem facing unions.

Ethan Kaplan's proposal exemplifies the strengths of empirical political economy, as practiced in economics departments. The evidence cited is all carefully identified from naturally occurring variation, and suggests a number of policies that could equalize political representation and increase turnout. Some of these suggestions highlight margins that are more likely to be thought of by an economist rather than a political scientist: for example the increased influence of money when media coverage of politics is low, suggesting that politicians, behaving somewhat rationally, trade-off responsiveness across pecuniary and popular constituencies.

A theme running through many of the essays is the power asymmetries that shape the functioning of our contemporary economy. Many economists dismiss the role of power because, as Naidu puts it in his essay, "under conditions of perfect competition and information, there is no scope for power." Talk about power is viewed as non-rigorous, or at least as belonging outside economics. But asymmetries between different groups abound: who has the upper hand in bargaining for wages and employment; who has market power and who gets to compete; who can move across borders and who is stuck at home; who can evade taxation and who cannot; who gets to set the agenda of trade agreements and who is excluded; who can vote and who is effectively disenfranchised. Some of these asymmetries are traditionally political imbalances; others are power imbalances that naturally occur in the market due to informational asymmetries or barriers to entry.

Policies that counter such asymmetries make sense not only from a distributional standpoint but also for improving aggregate economic performance. The policy essays tackle these asymmetries frontally, and suggest ways of rebalancing power for economic ends. Unions and wage boards can rein monopsony power in labor markets (Naidu and Dube); putting sand in the wheels of financial globalization can enhance the fiscal capacity

of the state (Zucman); regulating private finance can prevent crises (Admati and Mian); giving labor a greater say in trade agreements can improve the design of trade agreements (Rodrik); restricting campaign contributions and making it easier for poorer people to vote can increase the accountability of the political system (Kaplan).

Final words

The policy briefs that accompany this introduction range over a wide swathe of policy domains – social policy, taxation, labor markets, financial regulation, trade agreements, technology, and electoral rules. But their coverage is certainly not exhaustive; there are many important policy areas that remain untouched or are mentioned only briefly, and we have more contributions promised. The essays themselves are intended as first cuts, rather than definitive statements. We think of them as a modest beginning: a demonstration that mainstream economics produces relevant and imaginative policy ideas and an encouragement to other economists to contribute in the same vein. They are a proof-of-concept for the claim that economics can serve inclusive prosperity, and help build a society that is both fairer and does better job of living up to its productive potential.

Suresh Naidu is an Associate Professor of Economics at Columbia University. Contact: sn2430@columbia.edu

Dani Rodrik is the Ford Foundation Professor of International Political Economy at Harvard's John F. Kennedy School of Government. Contact: dani_rodrik@hks.harvard.edu

Gabriel Zucman is an Assistant Professor of Economics at the University of California, Berkeley. Contact: zucman@berkeley.edu

Endnotes

- 1 This is an introduction to the inaugural series of policy briefs prepared under the auspices of the Economics for Inclusive Prosperity (EfIP) network. The policy briefs and founding members of EfIP are listed at the end. We are grateful to the Hewlett Foundation for financial support. Special thanks to Margaret Levi and the Stanford Center for Advanced Study in Behavioral Sciences (CASBS) for stimulating the conversation that instigated this project.
- 2 A list of these initial policy briefs is at the end of this document.
- 3 We will not go into a detailed discussion of neoliberalism, a term that is commonly used by non-economists but not so much by economists. For standard treatments, see Brown (2006) and Harvey (2007). For a discussion from one of us, see Rodrik (2017).
- 4 There are many think tanks which rely on economists' ideas and engage them in thinking about policy issues. However, we are not aware of any *academic* network of economists focused on turning research and scholarship to policy use in the broad domain that we have called "inclusive prosperity."
- 5 Rodrik (2015) argues that the scientific nature of economics resides precisely in this ability to generate conditional hypotheses that can be confronted with evidence (even if not decisively tested).
- 6 Ash, Chen, and Naidu (2018) show that a teaching program on law and economics for judges, funded by a conservative donor, produced harsher prison sentences in criminal trials. Rodrik (2018) argues that investor and pharma lobbies distort the agenda of trade agreements towards clauses with high private gains but doubtful social benefits.

References

- Angrist, Joshua, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu, “Economic Research Evolves: Fields and Styles,” *American Economic Review, Papers & Proceedings*, 107(5), 2017.
- Ash, Elliott, Daniel L. Chen, and Suresh Naidu, “Ideas Have Consequences: The Impact of Law and Economics on American Justice,” unpublished paper, October 22, 2018.
- Brown, Wendy, “American Nightmare: Neoliberalism, Neoconservatism, and De-Democratization,” *Political Theory*, vol. 34, no. 6, December 2006.
- Diaz Alejandro, Carlos, “Trade Policies and Economic Development,” in Peter B. Kenen, ed., *International Trade and Finance: Frontiers for Research*, Cambridge University Press, 1975.
- Dixit, Avinash, “Governance Institutions and Economic Activity,” *American Economic Review*, 99:1, 2009.
- Harvey, David, *A Brief History of Neoliberalism*, Oxford University Press, 2007.
- Polanyi, Karl, *The Great Transformation: The Political and Economic Origins of our Time*, Beacon Press, Boston, 1944.
- Rodrik, Dani, *Economics Rules: The Rights and Wrongs of the Dismal Discipline*, W.W. Norton, 2015.
- Rodrik, Dani, “[Rescuing Economics from Neoliberalism](#),” *Boston Review*, November 6, 2017.
- Rodrik, Dani, “What Do Trade Agreements Really Do?” *Journal of Economic Perspectives*, Vol. 32, No. 2, Spring 2018.

Economics for Inclusive Prosperity (EfIP) network

Co-directors:

Suresh Naidu (Columbia)

Dani Rodrik (Harvard)

Gabriel Zucman (UC-Berkeley)

Founding members

Anat R. Admati (Stanford)

Sandra Black (Univ. of Texas)

Arindrajit Dube (UMass-Amherst)

Gauti Eggertsson (Brown)

Rebecca Henderson (Harvard)

Ethan Kaplan (Univ. of Maryland)

Anton Korinek (Univ. of Virginia)

Ilyana Kuziemko (Princeton) Atif Mian (Princeton)

Jesse Rothstein (UC-Berkeley)

Reed Walker (UC-Berkeley)

Note: University affiliations are for identification purposes only.

Inaugural EfIP Policy Briefs

<https://econfip.org/>

1. Suresh Naidu, Dani Rodrik, and Gabriel Zucman, “Economics for Inclusive Prosperity: An Introduction”
2. Anat R. Admati, “Towards a Better Financial System”
3. Sandra E. Black and Jesse Rothstein, “An Expanded View of Government’s Role”
4. Arindrajit Dube, “Using Wage Boards to Raise Pay”
5. Ethan Kaplan, “Election Law and Political Economy”
6. Anton Korinek, “Labor in the Age of Automation and Artificial Intelligence”
7. Suresh Naidu, “Worker Collective Action in the mid-21st Century Labor Market”
8. Atif Mian, “How to Think About Finance”
9. Dani Rodrik, “Towards a More Inclusive Globalization: An Anti-Social Dumping Scheme”
10. Gabriel Zucman, “Taxing Multinational Corporations in the 21st Century?”

Taxing multinational corporations in the 21st century

Gabriel Zucman¹

Globalization and the rise of intangible capital have increased tax avoidance opportunities for large firms dramatically. 40% of multinational profits are shifted to tax havens each year globally and the United States loses about 15% of its corporate income tax revenue because of this shifting. I discuss the evidence on the redistributive effects of international tax competition. I then present a proposal to reform the corporate tax that would remove any incentive for firms to shift profits or move real activity to low-tax places. Contrary to a widespread view, it is possible to tax multinational companies (potentially at high rates) in a globalized world, even in the absence of international policy coordination.

Introduction

There is a widespread view that taxing the winners from globalization—multinational companies, wealthy households, highly skilled individuals—is hard if not impossible in a globalized world. Companies can move abroad or shift profits to tax havens; the wealthy can move too or hide assets offshore.

The view that taxing multinational corporations is fraught with difficulties finds support in the undeniable reality of tax competition. Between 1985 and 2018, the global average statutory corporate tax rate has fallen by more than half, from 49% to 24%. In 2018, most spectacularly, the United States cut its rate from 35% to 21%. This cut is likely to exacerbate the race to the bottom for corporate tax rates throughout the world in the years to come. In September 2018, Theresa May pledged to cut the U.K. corporate tax rate to the lowest rate among G20 countries post-Brexit. A number of

other countries have announced their attention to further cut their rates (France, for instance, has planned to cut its rate from 33% today to 25% in 2022). Moreover, a large and growing fraction of profits are shifted to low-tax places. The prospects of taxing multinational companies at positive rates seem grim. Globally, some of the decline in corporate tax rates and loss of revenue caused by profit shifting has been compensated by base-broadening. But overall, the effective tax rates on corporate profits have declined a lot, almost in line with the decline in statutory tax rates (see Zucman, 2014). Moreover, in the United States the share of taxable corporate profits in GDP has fallen over time (due to the rise of the non-corporate business sectors and of tax-exempt corporations, known as S-corporations), reinforcing the decline in total corporate income tax revenue.

This essay argues that contrary to the widespread and intuitive view that corporate taxes are bound to fall, it

is perfectly possible to combine globalization and the taxation of multinational corporations—including at high rates. Not only is this possible, but it is also necessary to make globalization sustainable economically and politically. It seems indeed unlikely that globalization will continue to proceed if its main winners pay less in less in taxes, while those who don't benefit from it or are hurt by it—such as retirees, working-class individuals, and small businesses—have to pay more to make up for the lost tax revenue. For those who want to defend openness, it is critical to explain how globalization can concretely be combined with progressive taxation.

To contribute to this debate, I first discuss the evidence on the extent of corporate tax avoidance and the redistributive effects of international tax competition today. 40% of multinational profits are shifted to tax havens each year globally and the United States loses about 15% of its corporate income tax revenue because of this shifting. I then present a proposal to reform the corporate tax that would remove any incentive for firms to shift profits or move real activity to low-tax places. This reform would apportion the global, consolidated profits of firms proportionally to where they make their sales. Take a company that makes \$10 billion in profits globally and 20% of its worldwide sales in the United States. In the reform I describe, 20% of this company's global profits (i.e., \$2 billion) would be taxable in the United States. Such a sales-based apportionment would put an end to profit shifting as it exists today and dramatically alleviate the pressure towards lower corporate tax rates.

The redistributive effects of tax competition

How much do the various countries of the world win or lose in profits today because of tax competition? Tax competition between nations affects the location of profits in two ways. First, multinational companies have incentives to move tangible capital from high-tax countries to places where taxes are low. As emphasized in standard models of tax competition (see, e.g., Keen and Konrad, 2013), this relocation increases wages in low-tax places (to the extent the capital and labor are less than infinitely substitutable) and it can increase or decrease welfare in these countries. It also reduces wages and unambiguously decreases welfare in high-tax places. In contrast to international trade (which

can make losers within countries, but in standard theory enhances aggregate income in each country), tax competition makes some countries lose—and can even make most countries lose.

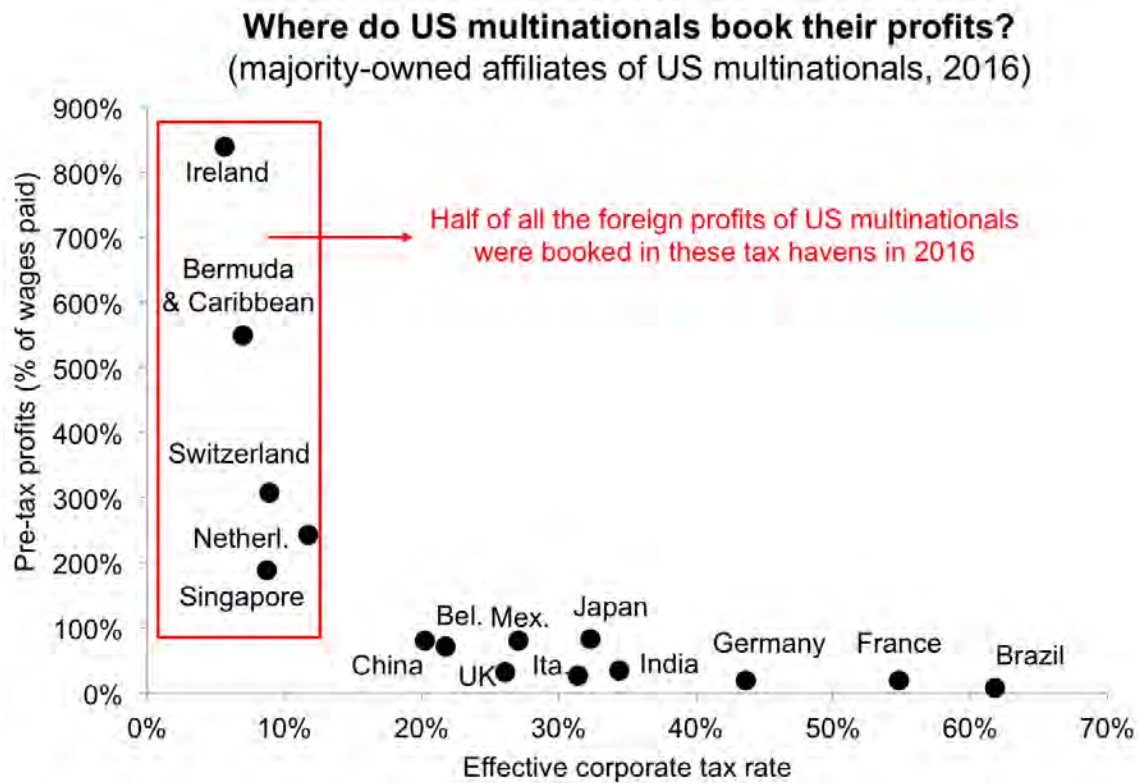
In addition to moving tangible capital to places where taxes are low, multinational companies shift paper profits to tax havens. They can do so in three ways: (i) by manipulating intra-group import and export prices (with affiliates in high-tax countries importing goods and services at high prices from related firms in low-tax countries), (ii) by using intra-group borrowing (with affiliates in low-tax places lending money to related firms in high-tax countries), (iii) by “locating” intangibles (such as patents, logos, algorithms, etc.) in tax haven subsidiaries.

Imagine that all countries had the same effective corporate income tax rate. That is, imagine there was perfect international tax coordination on both corporate tax rates and the definition of the tax base (same interest deduction and depreciation rules, for instance). In such a world, by how much would the profits booked by multinational companies in the United States rise compared to today's world? And by how much would they fall in low-tax places such as Ireland and Bermuda? There are two ways profits would adjust: some of the tangible capital located in low-tax places today would move back to high-tax places and profit shifting would disappear. To quantify the magnitude of these changes—that is, who wins and loses from tax competition—it is helpful to start by studying where multinationals book their profits today.

Profit shifting by U.S. multinationals

A vast literature studies profit shifting by U.S. multinationals, for one simple reason: the U.S. data are particularly good. The U.S. Bureau of Economic Analysis has a sophisticated statistical system to monitor its multinationals. A large sample of representative multinationals report detailed data annually to the Bureau since 1982; before that, benchmark surveys were conducted every five years. This dataset provides information about the foreign operations of U.S. multinationals abroad, including the profits booked in shell (or letter-box) companies in tax haven countries. Wright and Zucman (2018) use these data to study how much profits US multinationals have reported in each country and how much taxes they have paid abroad since 1966.²

Figure 1



Source: author's computations using the BEA survey of the foreign operations of U.S. multinationals for year 2016. Pre-tax profits are "profit-type returns" in Table II.F.1.; foreign corporate taxes paid are from Table II.D.1; the effective corporate tax rate is computed by dividing foreign corporate taxes paid by pre-tax profits.

The latest available data are for the year 2016. In that year, US multinationals made \$435 billion in profits abroad.³ Almost half of these profits were booked in just 5 tax havens: Ireland, Netherlands, Switzerland, Singapore, and Caribbean tax havens plus Bermuda. The amount of profits booked in Ireland alone (\$76.5 billion) exceeds the profits booked in China, Japan, Mexico, Germany, and France *together*. Taking into account the profits shifted to Puerto Rico (not covered by the BEA data) and other tax havens, the data show that about 55% of the foreign profits of U.S. multinationals are booked in tax havens today. The profits booked in haven affiliates are enormous compared to the wages paid by these affiliates. In haven affiliates, the ratio of pre-tax profits to wages is around 350% (for any dollar of wage paid, U.S. multinationals say they make 3.5 dollars in pre-tax profits). In non-haven affiliates, this ratio is below 50%. In other words, the location of profits has become dramatically disconnected from where firms employ workers and from they produce goods and services more broadly.

It is not hard to understand why U.S. multinationals book such a high fraction of their profits in Ireland, Netherlands, and similar places. As shown by the Figure below, these countries impose taxes on the profits of U.S. multinationals at very low rates—in a range of 5%-10%. That is, taxes paid by U.S. affiliates to these countries amount to 5%-10% of the profits booked in these countries. There is a strong correlation between where U.S. multinationals book their profits and the effective tax rate they face. For small countries, imposing low rates of around 5% to foreign profits is revenue-maximizing in the current international tax system: it allows them to attract a huge tax base, which generates large revenue when tax rates only marginally higher than 0 are applied.

Global profit shifting

U.S. multinationals are not the only ones to shift profits to tax havens. By drawing on foreign affiliates statistics

(similar to the BEA data) that were recently compiled in many economies, Tørsløv, Wier and Zucman (2018) estimate the extent of global profit shifting. They find that globally, 40% of multinational profits (profits made by corporations outside of the country where their parent company is incorporated) are booked in tax havens. This represents \$600 billions of profits which are made in high-tax countries each year, but end up being booked and taxed at very low rates in tax havens. Note that 40% of multinational profits shifted offshore, although a large figure, is lower than for U.S. multinationals alone (55%). That is, although multinationals from all over the world use tax havens, U.S. multinationals appear to use them particularly extensively.

The tax revenue costs of tax competition are sizable for many countries. Tørsløv, Wier and Zucman (2018) estimate that the United States loses about 15% of its corporate tax revenue because of the relocation of profits to low-tax places. Although tax havens do collect revenue on the huge bases they attract, profit shifting significantly reduces corporate income tax payments globally: for each \$1 paid in tax to a haven, close to 5\$ are avoided in high-tax countries. More than redistributing tax revenues across countries, profit shifting thus redistributes income to the benefit of the shareholders of multinational companies.

To better understand the high profits booked by multinationals in tax havens, it is helpful to decompose them into real effects (more tangible capital used by foreign firms in tax havens) and profit shifting effects (above-normal returns to capital and receipts of interest). This distinction matters because these two processes have different distributional implications. Movements of tangible capital across borders affect wages, since tangible capital has a finite elasticity of substitution with labor. By contrast, movements of paper profits (i.e., profit shifting) don't: for a given global profitability, whether income is booked in the United States or in Bermuda has no reason to affect workers' productivity in either of these places. Tørsløv, Wier and Zucman (2018) find that the high profits of multinationals in tax havens are mostly explained by shifting effects. Tangible capital is internationally mobile—and there is evidence that this mobility has become slightly more correlated with tax rates over the last twenty years. But globally, machines don't move massively to low-tax places; paper profits do. These offshore profits are not left dormant; they are either invested in global securities markets, or used for cross-

border mergers and acquisition, or used as collateral for loans that finance investments in the United States or other countries.

Of course, this finding does not tell us what would happen in a world without profit shifting, nor does it allow us to predict what will happen in the future. Logically speaking, it is entirely possible that multinationals will start moving more tangible assets to low-tax places if policy-makers reduce profit shifting opportunities but tax rates remain different across countries. What the data suggest is that, so far, profit shifting has swamped tax-driven capital mobility. But to address the fundamental challenges that globalization raises for corporate taxation, it is important to formulate reform proposals that would not only reduce profit shifting, but also reduce the incentives of firms to move real activity to low-tax places. Otherwise the risk is we may exacerbate forms of tax competition that are even more harmful than the competition for paper profits that is observed today.

Taxing multinationals in a globalized world

The good news is that there is a way to tax multinationals in a way that addresses both tax competition for real activity and for paper profits. At the country level, the corporate income tax base can be made largely inelastic by apportioning the global, consolidated profits of firms proportionally to where they make their sales. Concretely, if Apple sells 20% of its products in the United States, the U.S. federal government would say that 20% of Apple's global profits are taxable in the United States. This would put an end to profit shifting because firms cannot affect the location of their customers (they can't move their customers to Bermuda; and if they try to pretend that they make a disproportionate fraction of their sales to low-tax places, this form of tax avoidance is easy to detect and anti-abuse rules can be applied). This would also put an end to competition for real activity, because in such a system there is no incentive for firms to move capital or labor to low-tax places; the location of production becomes irrelevant for tax purposes.

U.S. states have successfully taxed companies operating in multiple states this way for decades, so this is a tried

and tested proposal. For instance, California, like all other States with a corporate tax, uses an apportionment formula to determine what fraction of corporate profits are taxable in California. Since 2013, apportionment is based on sales only: if a company makes 10% of its U.S. sales in California, then 10% of its U.S. profits are taxable in California (at a rate of 8.84% currently). Before 2013, the formula was more complicated, with apportionment based not only on the fraction of sales, but also the fraction of employment and tangible capital assets used in California by the corporation. Over time, the majority of U.S. States have gradually adopted apportionment formulas based mostly or only on sales. Other countries with sub-federal corporate taxes, such as Canada and Germany, use similar apportionment mechanisms. With globalization, countries are becoming more and more like local governments within a broader federation; and therefore using apportionment formulas like local governments have done for a long time is logical way forward.

Apportioning the global profits proportionally to where sales are made can be done unilaterally. International cooperation is always preferable, but because tax havens derive large benefits from tax competition, it is unlikely they will ever agree to meaningful changes to the international tax system (at least absent large economic sanctions). But any country is free to set its tax base as it sees fit; and not all countries need to use the same base (e.g., the same apportionment formula) for the corporate tax to work well. U.S. States have for a long time used different formulas (with some States such as Massachusetts apportioning profits based not only on the fraction of sales made in Massachusetts, but also the fraction of the capital stock in and wages paid in Massachusetts).

This reform of the corporate tax illustrates a simple yet powerful idea, namely that globalization and redistributive taxation are not incompatible. The corporate income is progressive, because although it is typically levied at a flat rate, equity ownership (and hence corporate profits) are very unequally distributed (see, e.g., Saez and Zucman, 2016 for evidence and equity—and more broadly wealth—concentration in the United States). The corporate tax reaches wealthy individuals regardless of whether profits are distributed to shareholders or retained within corporations. For example, although Facebook does not pay dividends, it pays corporate income tax (and would pay much more with sales apportionment of Facebook's profits),

and hence Mark Zuckerberg—Facebook's main shareholder—indirectly pays taxes this way.

Of course, the corporate tax does not necessarily entirely fall on shareholders. In principle, part of it may be shifted to labor. The incidence of capital taxes depends on the elasticity of capital supply, the elasticity of labor supply, and the elasticity of substitution between capital and labor. If both the labor and capital supply elasticities are small relative to the elasticity of substitution between capital and labor, then capital taxes (such as the corporate tax) fall on capital and labor taxes fall on labor. In a closed economy, it is unlikely that the supply of capital is very elastic. In an open economy, tax competition makes the supply of capital more elastic—and hence can contribute to shifting the incidence of the corporate tax to labor. The reform described here would annihilate tax competition and dramatically reduce the capital supply elasticity. The corporate tax would fall, like in a closed economy, mostly on capital. Raising the corporate rate to 35% (as was the case until 2017) or to 50% (as was the case in the 1950s, 1960, and 1970s) would significantly increase the effective tax rate on wealthy individuals in the United States, and the overall progressivity of the U.S. tax system. In turn, this would contribute to curbing the rise of inequality which has reached extreme levels in the United States compared to other developed economies (Alvaredo et al., 2018), although an exact quantification of this effect warrants further research. Beyond the effect on inequality, such a move would make the tax system fairer and hence more legitimate, eventually contributing to making globalization more sustainable in the 21st century.

Gabriel Zucman is an Assistant Professor of Economics at the University of California, Berkeley. Contact: zucman@berkeley.edu

Endnotes

- 1 This essay was prepared as part of the Economics for Inclusive Prosperity (EfIP) series of policy briefs.
- 2 The BEA data include two measure of profits: a financial accounting measure (“net income”) and economic measure (“profit-type return”). We use the economic measure, which in contrast to “net income” avoids double-counting of the profits of indirectly-held affiliates and excludes capital gains and losses. See the Appendix of Wright and Zucman (2018) for a detailed discussion.
- 3 This figure excludes profits booked in Puerto Rico (of the order of \$40 billion). Puerto Rico is a foreign country for US tax purposes (i.e., the federal corporate tax does not apply to profits booked in Puerto Rico) and is used extensively by US multinationals to avoid taxes.

References

- Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2018. *The World Inequality Report 2018*, Harvard University Press, <http://wir2018.wid.world>.
- Saez, Emmanuel and Gabriel Zucman. 2016. “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data,” *Quarterly Journal of Economics*, 131(2), 519–578.
- Tørsløv, L., Wier, L. and Zucman, G., 2018. “The Missing Profits of Nations”, NBER working paper No 24701.
- Wright, Thomas, and Gabriel Zucman. 2018. “The Exorbitant Tax Privilege”, NBER working paper #24983.
- Zucman, Gabriel. 2014. “Taxing Across Boarder: Tracking Personal Wealth and Corporate Profits”, *Journal of Economic Perspectives*, 28(4): 121–148.

Towards a More Inclusive Globalization: An Anti-Social Dumping Scheme

Dani Rodrik¹

Theory and empirics both suggest that international trade has sharp distributional implications. Furthermore, redistribution caused by trade is often viewed by the general public as more harmful or disruptive than other domestic market shocks. I discuss conditions under which there may be a legitimate case for restricting trade to promote domestic social inclusion, and propose a specific policy – a social safeguards clause – targeting those circumstances.

Trade and distribution

One of the remarkable implications of the theory of comparative advantage is that sharp distributional consequences are generically the flip side of the gains from trade. This point was first formalized in the famous Stolper-Samuelson (1944) theorem, which demonstrated that one of the factors of production would be left worse off in absolute terms as a consequence of opening up to trade. In a country where skilled labor is relatively abundant (compared to trade partners) and which has comparative advantage in skill-intensive goods, the loser would be unskilled labor. It is not simply that the gains from trade are distributed unevenly between skilled and unskilled labor; what is striking is that the losers suffer an absolute loss in real incomes.

The Stolper-Samuelson theorem is built on very specific assumptions: there are only two goods, two factors of production, and there is full mobility of factors between the two sectors of production. One might think that the stark distributional consequences it generates is a result of these specialized assumptions. In fact,

the result is remarkably robust and generalizes very broadly. Consider a world with any number of factors of production and any number of goods. Factors could be mobile, immobile, or anything in between. Suppose production takes place under neoclassical assumptions: that is, producers maximize profits and minimize costs using conventional production functions. Then, as long as a country does not fully specialize – i.e., as long as it continues to produce very close substitutes for importables – opening up to trade must leave at least one factor of production worse off in absolute terms.² The result that openness to trade creates losers is not a special case; it is the implication of a very large variety of trade models.

Nevertheless, until recently it was not uncommon to dismiss this as a theoretical result with little empirical support. Early research by trade economists looked for effects across the skill divide, and the effects there were not that large. Trade seemed to account for perhaps 10-20 percent of the rise in the skill premium. In retrospect, it appears that this work missed the scale of the distributional effects because they mostly focused on the wrong margins. More recent work has

focused on differences in labor markets across different communities and has uncovered much larger effects. Workers are apparently not very mobile spatially and communities that compete with imported goods can be hurt very badly by rising import competition.

Hakobyan and McLaren (2016) find that NAFTA produced modest effects for most U.S. workers, but an important minority suffered substantial income losses. Regions that were most affected by tariff reductions experienced significantly slower wage growth than regions that had no tariff protection against Mexico in the first place. The effect was greatest for blue-collar workers: a high-school dropout in heavily NAFTA-impacted locales had 8 percentage points slower wage growth over 1990-2000 compared to a similar worker not affected by NAFTA trade. The industry effect was even larger: wage growth in the most protected industries that lost their protection fell 17 percentage points relative to industries that were unprotected initially.³ These are very large effects, especially when one bears in mind that the net gains from trade generated by NAFTA apparently have been quite small, less than 0.5 percent at best (Romalis 2007, Caliendo and Parro, 2015).

In a well-known paper Autor, Dorn, and Hanson (2013) have documented the labor-market disruption caused by the “China trade shock,” which was not only large but also very persistent. These authors’ unit of analysis is the commuting zone. Their baseline result is that a commuting zone in the 75th percentile of exposure to Chinese import growth had a differential fall of 4.5 percent in the number of manufacturing employees and a 0.8 percentage point larger decline in mean log weekly earnings, compared to a commuting zone at the 25th percentile. They also find a significant impact on overall employment and labor force participation rates, indicating that this is an additional margin of adjustment to trade shocks. As the authors stress, this implies that the wage reductions are under-estimated, both because of increase in non-participation and the fact that the unemployed are more likely to have lower ability and earnings. Moreover, these local labor-market effects appear to have been highly persistent. The wage, labor-force participation, and unemployment consequences had not dissipated after a full decade of the China trade shock (Autor et al. 2016).

Trade, fairness, and appropriate remedies

How should such distributional effects be remedied? In a market economy, labor markets are buffeted constantly by shocks of different types. Jobs can be lost or displaced because of demand shocks, technology shocks, management decisions, and a host of other reasons. Trade is only one source of labor market disruption, and normally far from the most important one. Most economists would probably agree that there should be some kind of compensatory mechanism (unemployment and training benefits) when the shocks hit those at the bottom end of the labor market. They would also agree, however, that the safety net should not discriminate by the type of shock. If we are going to help those who are adversely affected by labor market disruptions, we should treat those who are hit by import competition differently from those who are displaced -- against their will -- for other reasons.

The view that policy makers should not be concerned by the nature of the underlying shock is predicated on an implicit judgement that all market shocks are alike and therefore require identical responses, if any. But this judgement is not consistent with basic moral intuitions. To make the point as starkly as possible, consider the following thought experiment. Suppose Harry and John run two firms that compete with each other. How do you feel about the following scenarios?

1. Harry works really hard, saves and invests a lot, comes up with new innovations, and outcompetes John, resulting in John and his employees losing their jobs.
2. Harry gets a competitive edge over John by finding a cheaper supplier in Germany.
3. Harry drives John out of business by outsourcing to a supplier in Myanmar, which employs workers in 12-hour a day shifts and under extremely hazardous conditions.
4. Harry brings workers from Myanmar to the U.S. under temporary contracts, and puts them to work under conditions that violate domestic labor, environmental, and safety laws.

These scenarios are isomorphic from a purely economic standpoint insofar as each creates losers as well as gainers in the process of expanding the overall size of the economic pie for the national economy. That is,

Harry's gains are larger than John's losses. They differ only in the manner in which these gains and losses are generated.

When I present these scenarios to my students, they react very differently to them, and I imagine most other audiences would too. Scenario 1 generally elicits the least opposition; what is happening seems to be the normal operation of a competitive market economy. Scenario 2 typically raises few concerns either – at least for an audience that is well educated and understands the benefits of international trade. But support drops sharply when I present scenarios 3 and 4. It appears there is something problematic with the exchanges described in the latter two scenarios. Some self-reflection may suggest that what is different is that these scenarios entail a form of market competition that would be considered unacceptable if it took place at home, and therefore has been legally ruled out in the domestic jurisdiction.

Some may remain unconvinced that the distributional burden created by scenario 3 is any different than that in scenarios 1 or 2. To economists, in particular, it may seem that the source of comparative advantage does not matter, even if it is abuse of labor rights. But then such economists should also be in favor of scenario 4 (which would of course break the law) – and I have met few who are willing to go that far. But then why should scenario 3 be OK if scenario 4 is not?

In recent work, Rafael di Tella and I carried out a survey in which we presented respondents with a news story about a possible factory closure that would leave hundreds of workers at risk of unemployment. Our “treatments” consisted of different explanations for why the factory might close. These included: a technological shock (automation), a demand shock (changing consumer preferences), management failures, and two trade shocks, outsourcing to a developed country (France) and outsourcing to a developing country (Cambodia). A control scenario where no specific shock is mentioned was also included. Then we asked two questions on how the government should respond: (a) whether the government should provide financial assistance to displaced workers, and (b) whether the government should restrict imports.

The results (shown in Figure 1) support three broad conclusions. First, respondents' willingness to provide financial compensation to workers is dependent on

whether the shock is trade related or not. Non-trade shocks increase willingness to provide financial support; trade shocks decrease it (in both cases relative to the control scenario). Second, trade shocks greatly increase preferences for import protection, relative to non-trade shocks. Third, there is a further difference between trade that involves a developed country and trade that involves a developing country. The preference for import protection is greatest in the case of outsourcing to a developing country.

Clearly our respondents draw sharp differences across the scenarios and how the government ought to respond. While financial compensation – safety nets – is viewed as appropriate for domestic market shocks, it is viewed unfavorably for trade shocks. And they viewed trade with developing countries as more problematic than trade with developed countries, exhibiting a preference for much greater import protection in the first case.

One way to interpret these results is through the lens of distributive fairness. International trade is viewed differently from domestic competition because certain kinds of international competition can undermine domestic norms with regards to what's an acceptable redistribution. (Note that a similar thing happens when competition from tax havens undermines the domestic tax regime, or when imports from jurisdictions with poor safety enforcement undermine domestic consumer safety rules.) This is the argument that corresponds to scenario 3 in the thought experiment above. In this case, compensation is generically inadequate because what is at stake is the surreptitious modification of the rules of the game – the undermining of domestic social bargains through the back door. Trade is not merely a market relationship, but an instrument for reconfiguring domestic institutions to the detriment of certain groups. One could argue that such instances require targeting directly the trade flows that have the alleged effect.

In summary, we need to distinguish between two different arguments for why trade may be problematic from a distributional – and hence social and political – perspective. When international trade operates just like any domestic form of market competition, it makes little sense to set it apart and decouple it from other approaches for dealing with inequality in labor markets at large (unemployment compensation, progressive tax systems, active labor market policies, employment-friendly macro policies, etc.). But when trade entails

practices that violate laws or norms embodied in our domestic institutional arrangements, and thereby undercuts domestic social bargains, it is legitimate to restrict the import flows that have the alleged effect.

In the specific context of trade with developing nations, what should be of particular concern for labor advocates is not low wages or labor costs per se, to the extent that those reflect labor productivity or alternative employment opportunities. Trade is unfair when competitive advantage is gained through the violation of worker rights in the exporting country. The proposed policy is a remedy against this kind of trade, to prevent social dumping.

A remedy against social dumping

A policy that targets social dumping must distinguish between true social dumping and regular market competition. Therefore it needs a domestic investigatory process of fact finding. To see how such a process can be devised we can take our cue from the prevailing trade remedy regime under the WTO. Two types of trade remedies are especially relevant: anti-dumping and safeguards.

The WTO allows countries to impose anti-dumping duties when imported goods are being sold below cost. In addition to determining dumping, domestic authorities must show a “material injury,” or threat thereof, to a domestic industry. And under the Agreement on Safeguards, countries are allowed a (temporary) increase in trade restrictions under a narrow set of conditions. Triggering the safeguards clause requires determination that increased imports “cause or threaten to cause serious injury to the domestic industry,” that causality from imports be firmly established, and that injury be not attributed to imports if there are multiple causes for it. Safeguards cannot be applied to developing-country exporters unless their share of imports of the product concerned is above a threshold. And affected exporters must be compensated by providing “equivalent concessions.”

A broader interpretation of safeguards would acknowledge that countries may legitimately wish to restrict trade for reasons going beyond competitive threats to the profitability of their industries.⁴ As I have discussed, distributional conflicts with domestic

norms or social arrangements are one such reason. We could imagine recasting the current agreement into an Agreement on Social Safeguards, permitting the application of safeguard measures under a broader range of circumstances. This would require replacing the “serious injury” test with another hurdle: the need to demonstrate broad domestic support, among all concerned parties, for the proposed safeguard measure.

To see how that might work in practice, consider what the current WTO agreement says:

“A Member may apply a safeguard measure only following an investigation by the competent authorities of that Member pursuant to procedures previously established and made public in consonance with Article X of the GATT 1994. This investigation shall include reasonable public notice to all interested parties and public hearings or other appropriate means in which importers, exporters and other interested parties could present evidence and their views, including the opportunity to respond to the presentations of other parties and to submit their views, inter alia, as to whether or not the application of a safeguard measure would be in the public interest. The competent authorities shall publish a report setting forth their findings and reasoned conclusions reached on all pertinent issues of fact and law.”

As written, the clause allows all relevant groups, and exporters and importers in particular, to make their views known, but it does not actually compel them to do so. Consequently, it creates a bias in the domestic investigative process towards the interests of import-competing groups, who are the petitioners for import relief and its obvious beneficiaries. This is also a key problem with hearings in anti-dumping proceedings, where testimony from other groups besides the import-competing industry is not allowed.

A key reform, then, would be to require the investigative process in each country to: (i) gather public testimony and views from all relevant parties, including consumer and public-interest groups, importers of the product(s) concerned, and exporters to the affected country, and (ii) determine whether there exists broad support among these groups for the application of the safeguard measure in question. Protectionism pure and simple would not have much chance of success if groups whose

incomes would be adversely affected by trade restrictions -- importers and exporters -- were necessarily part of the deliberative process and the investigative body had to determine whether these groups also support the safeguard measure. At the same time, when deeply and widely held social norms are at stake, these groups are unlikely to oppose safeguards in a public manner, as this would endanger their standing among the public at large. Imagine, for example, that forced labor was used in producing goods for export in country X, or that labor rights were widely and violently repressed. Exporters to country X and downstream users of X's products would find it difficult to publicly defend free trade with this country.

In less clear-cut cases, the main advantage of the proposed procedure is that it would force a public debate on the legitimacy of trade and when it may be appropriate to restrict it. It ensures that all sides would be heard. This is something which rarely happens. This procedure could also be complemented with a strengthened monitoring and surveillance role for the WTO, to ensure that domestic procedures are in compliance with the expanded safeguard clause. The specific oversight criteria might include transparency, accountability, inclusiveness, and evidence-based deliberation. An automatic sunset clause could ensure that trade restrictions do not become entrenched long after their perceived need has disappeared.

WTO panels would still have jurisdiction, but on procedural rather than substantive grounds. They would examine the degree to which requirements of democratic deliberation were fulfilled. Were the views of all relevant parties, including consumer and public-interest groups, importers and exporters, civil society organizations, sufficiently represented? Was all relevant evidence, scientific and economic, brought to bear on the final determination? Was there broad enough domestic support in favor of the opt-out or safeguard in question? The panels may rule against a country because the internal deliberations excluded an interested party or relevant scientific evidence. But they would not be able to rule on the substantive claim—whether in fact the safeguard measure serves the public interest at home by furthering a domestic social purpose. This echoes the procedural emphasis in the existing Agreement on Safeguards, although it greatly increases the scope of its application.

The proposed procedure would force a deeper and

more representative public debate on the legitimacy of trade rules and on the conditions under which it may be appropriate to suspend them. The most reliable guarantee against abuse of opt-outs is informed deliberation at the national level. The requirements that groups whose incomes would be adversely affected by the opt-out—importers and exporters—participate in the deliberations and that the domestic process balance the competing interests in a transparent manner would minimize the risk of protectionist measures benefiting a small segment of industry at large cost to society. A safety valve that allows principled objections to free trade prevail makes it easier to repress protectionist steam.

Even though domestic interests would presumably dominate the deliberations, the consequences for foreign countries need not be entirely overlooked. When social safeguards pose serious threat to poor countries, for example, non-governmental organizations and other groups may mobilize against the proposed opt-out, and those considerations may well outweigh ultimately the costs of domestic dislocations. A labor union may win protection when its members are forced to compete against workers abroad who toil in blatantly exploitative conditions. They are much less likely to carry the day against countervailing domestic interests when foreign working conditions reflect poor productivity rather than repression of rights. As the legal scholar Robert Howse notes, enhancing confidence in the ability of domestic deliberations to distinguish between legitimate domestic regulations and protectionist “cheating” should allay concern that domestic measures are purely protectionist. “Requiring that regulations be defensible in a rational, deliberative public process of justification may well enhance such confidence, while at the very same time serving, not frustrating, democracy” (Howse 2000, p. 2357). The proposed safeguard would be the embodiment of the principle that countries have the right to uphold national standards when trade undermines broadly popular domestic practices, by withholding market access or suspending WTO obligations if necessary.

Current safeguard procedures require most-favored nation (MFN) treatment of exports, permit only temporary measures, and demand compensation from the country applying the safeguard. These need to be rethought in the context of the broader arrangement I am proposing. MFN treatment will often not make sense. If the safeguard is a reaction to labor abuses in a

particular country, it is appropriate to direct the measure solely against imports from that country. Similarly, an ongoing abuse will require ongoing use of the safeguard. Instead of imposing temporary relief, it would be better to require periodic review or a sunset clause that could be revoked in case the problem continues. This way trade restrictions or regulations that hamper other countries' interests are less likely to become ossified.

The issue of compensation is trickier. When a country adopts a safeguard measure, the logic goes, it revokes a "trade concession" it had previously granted to other countries in an internationally binding agreement. Those other countries are entitled to receive equivalent concessions or to revoke some of their own concessions in return. In a dynamic world with near constant change, the nature of the concessions that a country grants to others cannot be predicted perfectly. This uncertainty turns international trade agreements into "incomplete contracts." When unforeseen developments change the value or cost of trade flows—because of new technologies (genetic engineering), say, or new values (on the environment), or new understandings (on desirable development strategy)—who controls rights over those flows? The requirement of compensation places those rights squarely with the international trade regime; the exporter can continue to demand market access on the original terms. But we might just as legitimately argue that the value of the original concessions depend on the circumstances under which they were provided. Under this interpretation, an exporter could not claim a benefit that did not exist, nor the importer be forced to suffer a loss that was not originally contemplated, when the agreement was signed. This would bring control rights closer to nation states and sharply limit the amount of compensation that exporters could expect.

Authoritarian regimes likely will become easier targets for safeguard action by democratic nations when their exports cause problems in those nations. Even though some of their labor practices, for example, will be easy to justify, others may not be. Minimum wages significantly lower than in rich countries can be rationalized in the domestic debate by pointing to lower labor productivity and living standards. Lax child labor regulations are often justified by the argument that it is not feasible or desirable to withdraw young workers from the labor force in a country with widespread poverty. In other cases, arguments like these carry less weight. Basic labor rights such as non-discrimination, freedom of association, collective bargaining, and prohibition

of forced labor do not cost anything. Compliance with these rights does not harm, and indeed possibly benefits, economic development. Gross violations constitute exploitation of labor, and will open the door for safeguards in importing countries on the ground that they generate unfair distributional costs.

Broadening safeguard action in this manner would not be without its risks. The possibility that the new procedures are abused for protectionist ends and open the door to unilateral action on a broad front, despite the high threshold envisaged here, has to be taken into account. But as we have already seen with the rise of Trump, doing nothing is not riskless either. Absent creative thinking and novel institutional designs, the tensions created by globalization will reinforce the protectionist backlash. That would be far worse than the safeguard regime I have just described. Moreover, qualms about the protectionist slippery slope have to be tempered by considering the abuse that occurs under the existing rules, without great detriment to the system. If mechanisms with explicit protectionist intent, such as anti-dumping, have not destroyed the multilateral trade regime thus far, it is not clear why well-designed exit clauses would have consequences that are worse.

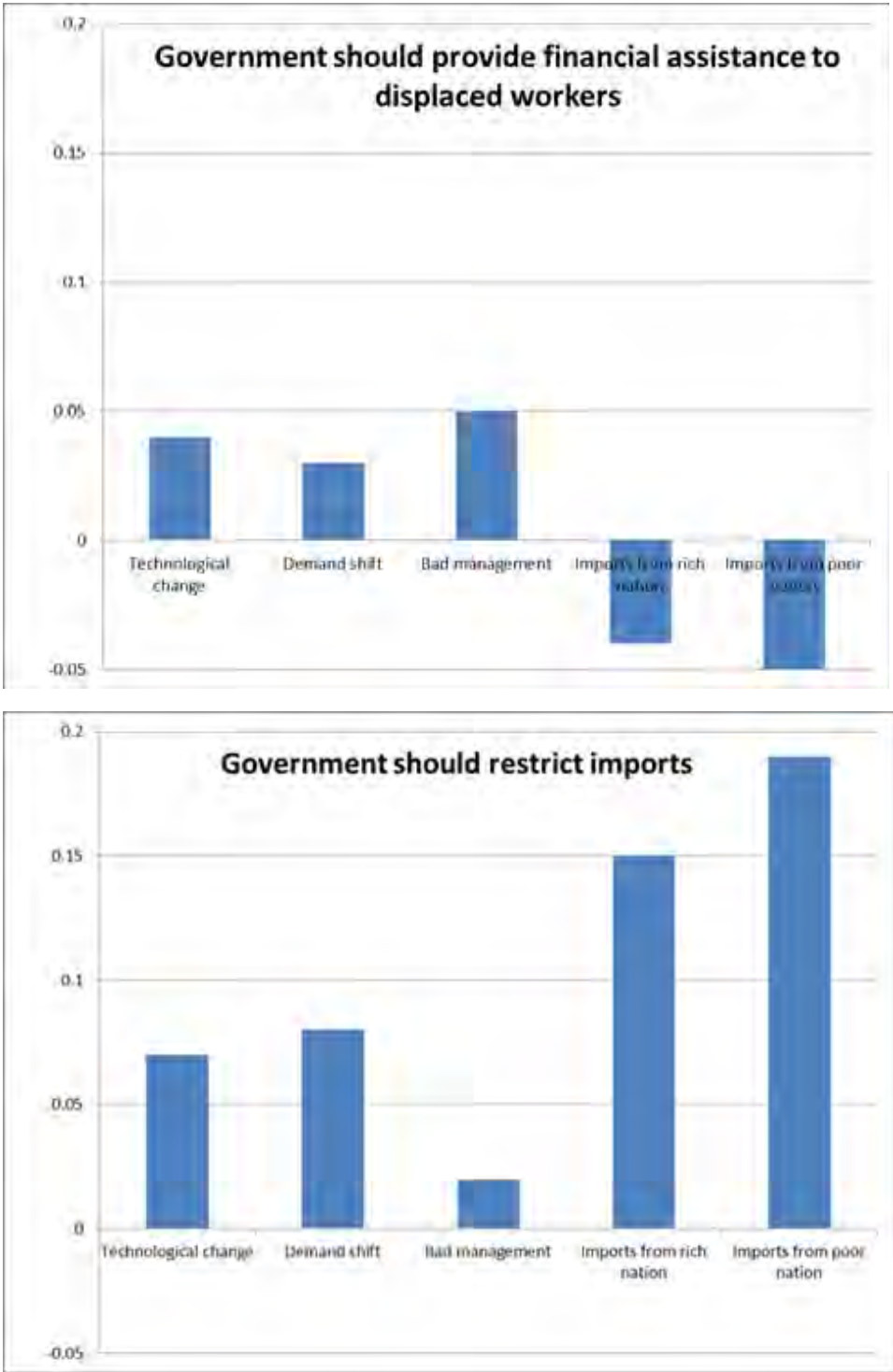
I address two remaining questions briefly. First, would such a scheme affect developing countries and their development prospects adversely? I would argue not, since the aim of the proposal is to delegitimize unwarranted protectionism (against developing countries in general) by enabling trade restrictions in those, relatively narrow range of circumstances where they are warranted on social grounds. My hypothesis is that generalized protectionism is rendered more likely in the absence of such a clause against social dumping. Moreover, the social safeguards described here could be paired with a "development box" that provides developing countries their own, enhanced policy space with respect to the use of industrial policies (Rodrik 2011). Such an exchange of policy space between the advanced and developing nations would benefit both partners, without necessarily harming prospects for global trade.

Second, why not address labor rights by incorporating labor clauses directly into international trade agreements, instead of providing domestic safeguards?²⁵ This has been the preferred route for two decades now, but with very meager results (Rodrik 2018b). Experience with aid conditionality shows that trying

to get countries to change their policies in return for continued material benefits (financial assistance or continued market access) does not have a very good track record. Regardless of what happens in trade agreements, there needs to be a domestic mechanism to address the problems discussed in this essay.

Dani Rodrik is the Ford Foundation Professor of International Political Economy at Harvard’s John F. Kennedy School of Government. Contact: dani_rodrik@hks.harvard.edu

Figure 1 Marginal effects on shares of respondents that respond favorably to statement on the chart (relative to control)



Endnotes

- 1 This essay was prepared as part of the Economics for Inclusive Prosperity (EfIP) series of policy briefs.
- 2 The proof of this result is sketched in Rodrik (2018a). Let the unit cost of production for the importable sector that is being liberalized be expressed as $c = \varphi(w_1, w_2, \dots, w_n)$, with w_i denoting the return to the i^{th} factor of production used in that sector. Since payments made to the factors must exhaust the cost of production, changes in unit costs are a weighted average of changes in payments to each of the factors, where the weights (in perfect competition) are the cost shares of each factor. In other words, $\hat{c} = \sum \theta_i \hat{w}_i$, where a “hat” denotes proportional changes, θ_i is the cost share of factor i , and $\sum \theta_i = 1$. Consider what happens with trade liberalization. The effect of trade liberalization is to raise the domestic price of exportables relative to importables. Let the importable described above be the numeraire, with price fixed at unity. We are interested in what happens to the returns of factors used in the importable. Since this good is the numeraire, we have the equilibrium condition $c = \varphi(w_1, w_2, \dots, w_n) = 1$, stating equality between price and unit cost (the zero-profit condition). As long as the good continues to be produced, this condition holds both before and after the liberalization. Therefore $\sum \theta_i \hat{w}_i = 0$. Hence there must be at least one factor of production, call it the k^{th} factor, such that $\hat{w}_k \leq 0$. (The inequality will be strict when goods differ in their factor intensities.) Meanwhile exportable prices have increased ($\hat{p} > 0$), thanks to the liberalization. Hence, $\hat{w}_k \leq 0 < \hat{p}$ and the return to the k^{th} factor declines in terms of both the importable and exportables, producing an unambiguous fall in real returns, regardless of the budget shares of the two goods.
- 3 Given the nature of identification, these results refer to relative wage changes. But since real wages of blue collar workers have generally stagnated, they are telling also about the magnitude of real income impacts.
- 4 I have written about the social safeguards clause in Rodrik (1997 and 2011), on which the following paragraphs are based. Legal aspects are discussed in Shaffer (2019).
- 5 See Tucker (2018) for a recent such proposal. Tucker advocates an international agreement that explicitly targets higher unionization rates, allowing countries discretion on how to get there. These targets would be paired by international arbitration that can be initiated by labor groups.

References

- Autor, David, David Dorn, and Gordon Hanson, “The China Syndrome: Local Labor Market Effects of Import Competition in the United States,” *American Economic Review*, 103(6), 2013, 2121–2168.
- Autor, David, David Dorn, and Gordon H. Hanson, “The China Shock: Learning from Labor Market Adjustment to Large Changes in Trade,” *Annual Review of Economics*, 8, 2016, 205–240.
- Caliendo, Lorenzo, and Fernando Parro, “Estimates of the Trade and Welfare Effects of NAFTA,” *Review of Economic Studies*, 82, 2015, 1–44.
- di Tella, Rafael, and Dani Rodrik, “Labor Market Shocks and the Demand for Trade Protection: Evidence from Online Surveys,” 2019, forthcoming.
- Hakobyan, Shushanik, and John McLaren, “Looking for Local Labor Market Effects of NAFTA,” *Review of Economics and Statistics*, 98(4), October 2016, 728–741.
- Howse, Robert, “Democracy, Science, and Free Trade: Risk Regulation on Trial at the World Trade Organization,” *Michigan Law Review*, Vol. 98, No. 7, June 2000.
- Rodrik, Dani, *Has Globalization Gone Too Far?* Institute for International Economics, Washington, DC, 1997.
- Rodrik, Dani, *The Globalization Paradox: Democracy and the Future of the World Economy*, W.W. Norton, New York, 2011.
- Rodrik, Dani, “Populism and the Economics of Globalization,” *Journal of International Business Policy*, vol. 1, 2018(a).
- Rodrik, Dani, “Can Trade Agreements Be a Friend to Labor?” *Project Syndicate*, September 14, 2018 <https://www.project-syndicate.org/commentary/trade-agreement-labor-provisions-small-practical-effect-by-dani-rodrik-2018-09>.
- Romalis, John, “NAFTA’S and CUSFTA’S Impact on International Trade,” *Review of Economics and Statistics*, 89(3), August 2007, 416–435.
- Shaffer, Gregory, “Reconceiving Trade Agreements to Address Social Inclusion,” *Illinois Law Review*, 2019, forthcoming.
- Stolper, Wolfgang, and Paul A. Samuelson, “Protection and Real Wages,” *Review of Economic Studies*, 9, 1941, 58–73.
- Tucker, Todd N. “Seven Strategies to Rebuild Worker Power for the 21st Century Global Economy: A Comparative and Historical Framework for Policy Action,” Roosevelt Institute, September 2018.

Worker Collective Action in the 21th Century Labor Market

Suresh Naidu

Private sector union density in the United States has fallen below 7%, but new historical evidence shows high union density played an important role in compressing the US income distribution at mid-century and lowering intergenerational income persistence. Other recent evidence on pervasive labor market power suggests that unions may be able raise wages without severe dis-employment effects, and may alleviate inefficient contracting problems. Despite substantial survey evidence indicating latent demand for unions, employers have successfully fought unionization efforts in rising service sectors, and a combination of legal restrictions and economic transformations have impaired the ability of US unions to solve collective action problems at the appropriate scale – an issue that economics may be able to help ameliorate.

What do unions do? Recent economic evidence

Private sector union density in the United States is a terribly low 7%, and even in its last, institutionally idiosyncratic redoubt – among public sector workers – the labor movement has recently been greatly weakened by the *Janus*¹ decision. Despite these blows, organizing collective action via the labor market remains a political economy lever that can't be ignored. Unions both address pervasive labor market failures and increase the political voice of the bottom half of the income distribution. Historically, unions were an institution that accomplished three objectives: economic redistribution via higher wages for unskilled workers, better workplace amenities and allocations of control rights inside the firm, and political representation.

Farber et al. (2018) use Gallup data to examine patterns of unionization and inequality in the 20th century. They find that union density over the 20th century correlated with negative selection into unions even as the union income premium and the relatively more compressed within-union income distribution stayed relatively constant. Expanding union density lowers inequality by compressing wages among union members, as well as by increasing wages of lower-skilled workers. This mechanism explains correlations between union density and income inequality in both the time-series and state-year fixed effects specifications.² Figure 1 presents evidence that points to union membership virtually eliminating the correlation between father's income and own income, suggesting unions do a lot for intergenerational mobility as well.

But alongside this more descriptive evidence, more causal evidence paints a very different picture, one

that seems difficult to reconcile with the stable union premium. Beginning with Dinardo and Lee (2004), and continuing with Lee and Mas (2012) and Frandsen (2014), economists have looked at the differences between close wins and close losses in NLRB elections, and found surprising effects: little effect on firm survival and profits, but also little effect on wages. What effects there are seem to be partly about composition (high-skilled workers, including managers, leave and low-skilled workers come to union jobs).

But union elections only impose the “duty to bargain in good faith”; only 60% of union recognition wins turn into first contracts after 2 years. In reality the evidence from these NLRB studies highlights the other little-known fact about labor law: at least since the Supreme Court’s 1930s decisions, and certainly since the 1947 Taft-Hartley bill, the formal NLRB architecture has been more about weakening worker collective action

than strengthening it. The number of workers added to unions via NLRB elections is tiny (Figure 2), and most unionized workers are joining already-unionized firms, not unionizing the firms they are already in. Frandsen (2017) shows that there is bunching at close elections, and the asymmetry in the bunching varies depending on whether Republicans or Democrats are in control of the NLRB. When Republicans are in charge, there are a suspiciously high number of close union losses, suggesting employers can fight harder without being sanctioned. And fight they do, as the firing rate of pro-union workers shows in Figure 4. Among labor lawyers, playing by NLRB rules is widely acknowledged to be a losing strategy. The bottom line: The NLRB certification process does not regularly result in an increase in union power, and this is particularly true in close elections.

So if union certification by the federal government doesn’t increase union power, what does?

Figure 1 Survey evidence on rank-rank coefficient (IGE) between father and son’s household income and its interaction with son’s union status. Data sources are American National Election Survey, the 1973 Occupational Changes in a Generation Survey, and the General Social Survey. From Jacome, Kuziemko, and Naidu (2018)

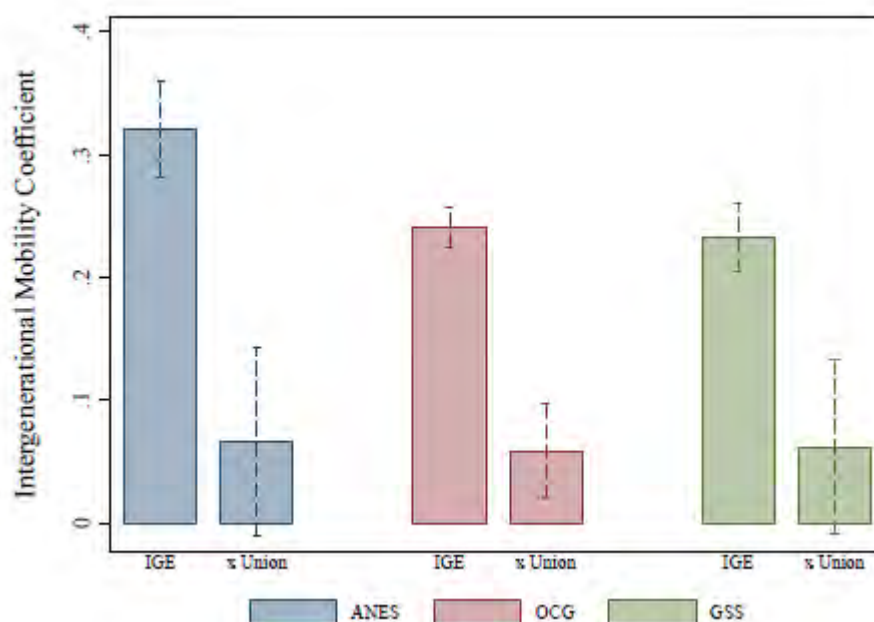


Figure 2 Number of employees in NLRB elections

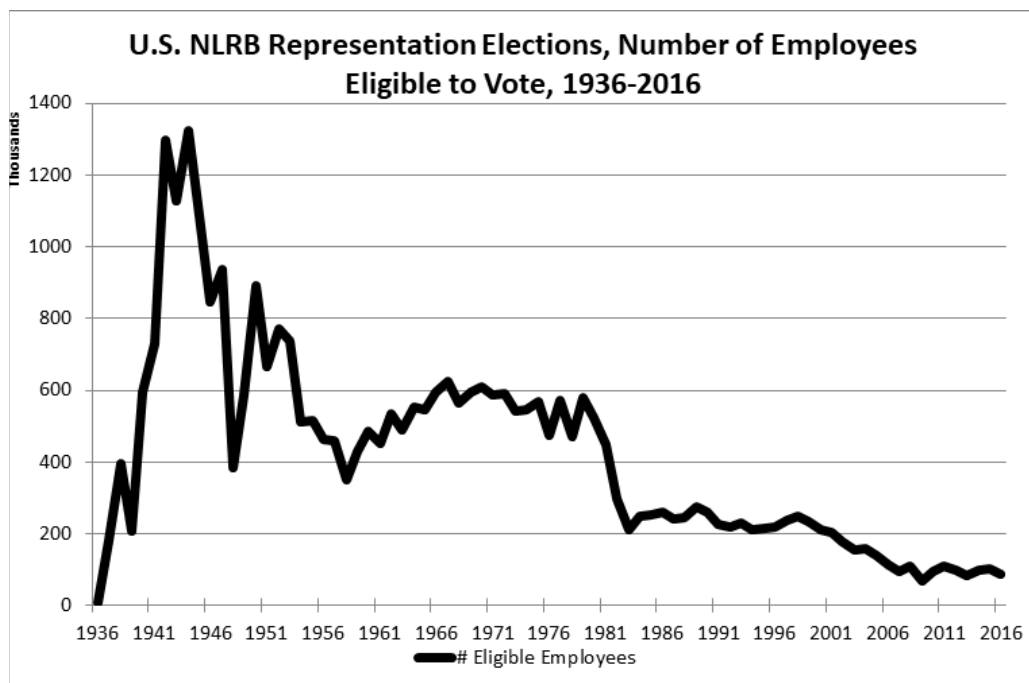


Figure 3 Asymmetrically close losses in NLRB elections when Republicans control the NLRB. From Frandsen 2017

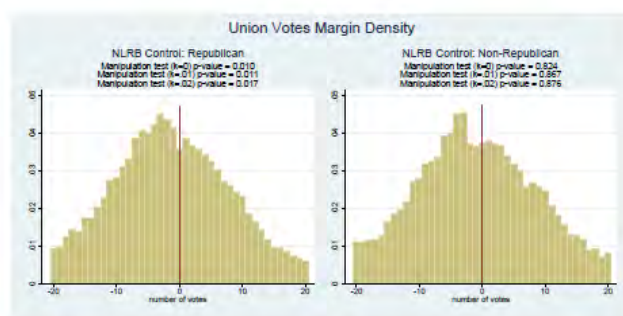
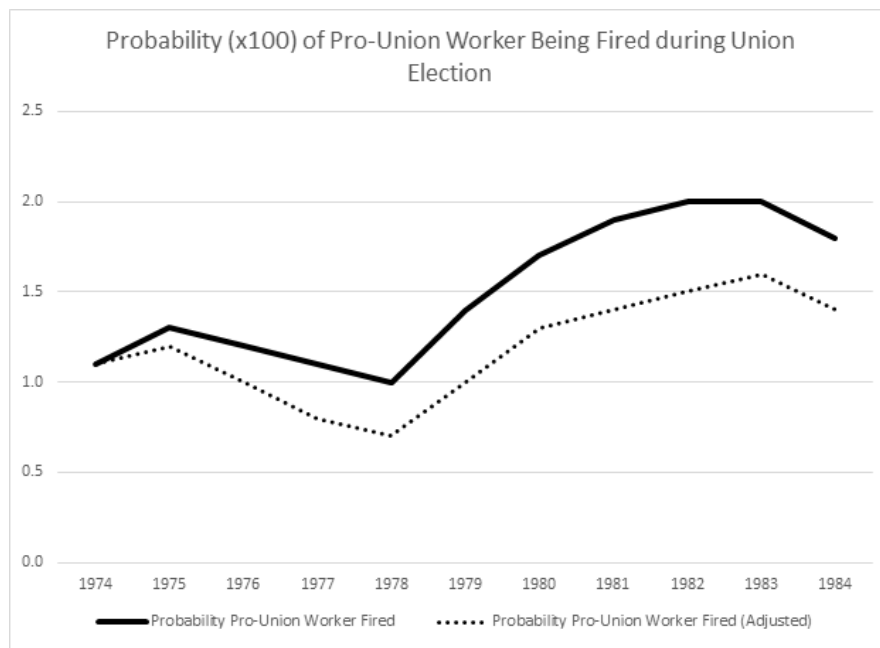


Figure 12: Density of the union margin of victory in terms of number of votes for elections with at least 20 voters. The left panel is for elections held during periods when Republicans controlled the NLRB. The right panel is for elections held during periods when Non-Republicans (i.e., Democrats and independents) controlled the NLRB. Data are from the NLRB. Histogram reproduced from Frandsen (2015).

Figure 4 Firing rate of pro-union workers. From Schmitt and Zipperer (2007)



Market power makes union power efficient

The inequality of bargaining power between employees who do not possess full freedom of association or actual liberty of contract and employers who are organized in the corporate or other forms of ownership association substantially burdens and affects the flow of commerce.
—National Labor Relations Act (1935)

Before tackling the question of union power, it is worth considering what “power” can mean in economics more broadly. “Power is important” is taken as axiomatic by many non-economists, but economists have a reflexive rebuttal: under conditions of perfect competition and information, there is no scope for power. Indeed, an old Samuelson-ism argued that it does not matter if labor hires capital or capital hires labor, and many economists think the term “power” is not rigorous nor a properly economic idea.

But labor movements and the economists closest to them have always had compelling counter-narratives about why the boss had the whip hand in the labor transaction.³ Institutionalist labor economists such as Sumner Slichter, John Dunlop, and Lloyd Reynolds all believed that frictions were pervasive in the labor market. One form of friction is imperfect mobility,

which means that, from the perspective of workers, jobs are imperfect substitutes. This lack of mobility could be either due to few employers in a given skill-location segment, costly job search, or non-wage differentiation. Employers set wages to take advantage of this, losing a few workers in order to depress wages for the ones that remain.

What unions do makes a lot of sense in models with monopsony. Monopsony implies that unions can a) raise the wage within limits without necessarily costing jobs⁴ and b) replace the individual labor-supply curve facing the firm with a much more efficient bilateral bargain. More broadly, monopsony means that the labor market interventions become the site of economic redistribution, in addition to (or instead of) the tax code, and so politically organized workers become an important constituency for redistribution via the labor market.

Beyond power over the wage, the default rule in the employment relationship is that employers have the right to command workers on the shop floor. This results in plenty of inefficiently allocated control rights, as there are many workplace decisions where workers have superior information about their cost of doing things. A union contract can reallocate these decision rights toward the efficient division, and evidence in Ash, Macleod, and Naidu (2018) suggests that this reduces

labor conflict (measured by strikes). Union contracts are efficiency-enhancing workplace constitutions.

Other canonical models of the labor market have a variety of notions of power explicitly built into them. Diamond-Mortensen-Pissarides search models use Nash bargaining between individual employers and workers, Stole-Zweibel has bargaining when workers can exit individually, Shapiro-Stiglitz has efficiency wages, where employed workers get rents so that the threat of unemployment secures effort provision. Sociologists and political scientists often discuss power; this idea has been integrated into a variety of economic models of the labor market, and recent evidence shows that outside options are generically important for wage determination (Caldwell and Harmon 2019, Caldwell and Danieli 2019). Monopsony is simply one particular variant of economic power, and one that empirically can explain many facts in the low-wage segment of the labor market.

Why have unions declined? Economic and political forces

There are of course the usual suspects; globalization, technological change, increased skills of workers/flexibility of firms. Union commitments push firms out of business (although even more onerous bondholder commitments are sacrosanct). But there are some anomalies: Canadian union density remains double the US level, despite superficially similar institutions and shocks⁵. Ghent system countries retain higher union density, due to unemployment benefits being administered by big cross-industry unions. The decline in unionization is not confined to tradable manufacturing; construction and transportation have seen similar declines.

Further, it is not clear exactly why services are so difficult to organize. They are not subject to international competition, and they are readily accessible to organizers (unlike isolated factories). One argument is that they have low barriers to entry, and so there are no rents to capture; but then these firms would also not be terribly profitable. Another possible story is that suppliers of other inputs or factors (e.g., landowners or financiers or upstream producers) extract most of the profit, and so low-skilled work is at the low-margin end of the value chain.

Another oft-heard agreement is that there was something unique about the kinds of industries that unions were strong in, with high fixed capital and strong complementarities of firm-specific skills across workers making strikes quite easy to win. I return to this below.

While these explanations can account for the almost universal decline in unions across the OECD, a leading candidate for the peculiarly rapid deunionization in the United States is employer and government opposition to unions, shown in Figure 3 above. Some of this can be overcome with policy alone. Survey evidence reveals that workers want to join unions and there is significant latent demand for unionization (particularly for the selective benefits such as health care that unions provide). But organizational capacity to take advantage of worker demand at the scale necessary to extract rents still needs to be built.

What can be done? And can economics help?

At the end of the day, what unions do is organize collective action on the basis of work. From strikes to pickets to phone banks to grievances, unions are powerful because they leverage the common interests that workers in a firm, occupation, or industry have into bargaining power and political power. It may very well be that firm-specific unions are artifacts of a particular technological period, or only feasible when big firms are also employers of low-wage labor, or in economies relatively insulated from trade with large, developing countries. But to the extent that unions are the outcome of a conflict between employers (or employers' employers) and collectively organized workers, measures that raise the capacity of unions to solve their collective action problems will, in theory, raise union efficacy, and likely density and coverage.⁶

Before going too far down this path, it is worth being realistic: Unions will not return to their midcentury density without truly radical policy and organizational changes. In the 1930s and 1940s the mobilizations of the CIO followed by the National War Labor Board essentially made union membership the “default option” in the key sectors of American industry. Any comparable change today would have to move 70 million workers into unions within a decade and a half. Despite this formidable outlook, it is still worth

considering what types of organizations can leverage work in the 21st century as a locus of shared identity and collective action.

Further, two economic trends might push in the direction of easing worker collective action. The first of these is the rise of platforms and large employers for low-skilled work, which are natural monopsonies but also make it easier to coordinate activities of workers in a sector. As the 19th century factory brought craftsmen together under one roof to reap productivity gains, it also allowed once-dispersed suppliers of labor to organize themselves collectively.

An example of this is the portable benefits platform for low-wage high-turnover workers being piloted by some labor organizations. By organizing workers to use a platform for a concrete service with increasing returns, it also gives the capacity for collective action, and regulation can demand that platforms must contract with *some* portable benefits platform. Further, the organization can use the platform to directly compete with the other platforms, forming a combination “strike-and-worker-cooperative” that can amplify collective bargaining power. Worker ownership of a potential competitor platform makes the threat of withdrawing labor from other platforms much more credible and costly.

A second force is the rise of personalized service work, where workers and customers meet in spaces not policed by managers (e.g., home health care workers, various retail workers, delivery workers, etc.). The traditional organizing “salt” strategy of having to get a few dedicated organizers employed as workers in highly monitored, private spaces might give way to a more “swarm” based strategy, as workers are organized via their many interactions with pro-union customers. On the flip side, however, customers may be more likely to blame the workers and the union for poor service, making it harder to build customer-worker alliances (Naidu and Reich 2018).

There are of course a variety of policy options that could encourage unionization on the margin, from exempting unionized firms from other labor regulation, facilitating union recognition (e.g., card check and Taft-Hartley repeal) and minority unionism, and institutionalizing large-scale worker organizations as distributors of benefits such as unemployment insurance (as in the Ghent system) or health care (some US unions run

extremely efficient nonprofit health care insurance, and may provide a scaffolding for broader health coverage).

But what I find most interesting is thinking about unions as organizations that can be made to work more effectively by deploying economics. It is surely a much smaller lift to use the tools of economics to restore collective power to workers in the advanced countries than to alleviate problems of international development! Indeed, unions are increasingly sophisticated, data-intensive organizations. My first glimpse of the potential of union data for collective action came after the 2012 election, where I noticed the Obama campaign’s sophisticated use of data and field experiments, and knew that the AFL-CIO’s electoral arm had been pivotal in building this operation, but that it had not been deployed for labor organizing. Since then I have worked with a number of labor organizations of varying sophistication and size, and the key thread unifying the problems was facilitating a variety of types of collective action, from meeting attendance to political contributions. Unions understand that their success depends on getting their members (or potential members) to operate in concert.

Lowering the costs of collective action

The canonical Olsonian analysis of unions argues that unions can’t survive without compulsion, because the public good of the collective bargaining agreement is vulnerable to the free-rider problem. However, economists have learned a lot about how humans cooperate in the wild, which suggests that the free-rider problem (dues or political contributions, picketing and strike compliance, or simple participation in union activities) is not insurmountable.

The first place to look is selective benefits. What do workers get by being in the union that they do not get otherwise? In right-to-work states, where workers can opt out of union dues, the answer is often – perhaps surprisingly – training programs. In focus groups I observed with a large NYC local, training programs were the union benefit workers were most enthusiastic about. This is backed up by a recent experiment by Hertel-Fernandez (2018), who randomizes messages sent to the members of an Iowa teachers’ union: Members who were reminded of the training programs

were the only treatment arm who differentially were likely to vote against a decertification.

Another idea is to use what we know from behavioral economics to make pro-sociality salient for (potential) union members. A large literature has documented pro-social preferences in a variety of public goods environments, as well as mediators that make pro-sociality more or less expressed. One important insight from behavioral economics is that a sizeable fraction of agents exhibit strong reciprocity: They punish free-riders even at a cost to themselves. The get-out-the-vote literature (see Gerber and Green 2017 for a survey) has leveraged numerous insights from social psychology and behavioral economics to move voter turnout, and these tools might be even more effective in the workplace, where the competition among information sources is less severe.⁷

Recent research has also shown the importance of networks and information diffusion in facilitating collective action, including strategic complements (attending membership meetings) or strategic substitutes (e.g., pickets and political contributions). Gonzalez (2018) shows that protest attendance increases among high school students when their peer groups of friends from junior high are more likely to go, with the critical threshold being around 40%, suggesting strategic complementarities. In contrast, Cantoni et al. (2018) have experimental evidence showing that Hong Kong student protest attendance looks much more like strategic substitutes, where students protest less the more they think other students are going to attend. This result was replicated among German party activists by Hensel et al. (2018), with activists less likely to knock on doors when told that a greater-than-expected number of activists were going to knock on doors. A way to reconcile these findings is to take the network model of Ballester, Calvo-Armengol, and Zenou (BCZ 2006) and note that collective action is likely complementary in network ties (e.g., actions are strategic complements in friendship networks) but substitutes globally (i.e., in the whole population). Encouraging collective action would take the form of solving coordination games within cliques of associates and friends, but providing incentives to overcome the free rider problem in the whole population. The BCZ paper tells us how to find the worker(s) that have the largest impact on collective action (those with highest intercentrality in the network), and this corresponds to the informal wisdom among organizers: Find the most

prestigious worker on a shop floor and convince them first. Mapping intrafirm networks across workers may allow this insight to be used as a predictive heuristic for prioritizing organizer efforts.

Finally, and most economist-friendly, is the possibility of applying ideas from mechanism design to the solution of labor's various collective action problems. Depending on the relative strength of strategic complements versus substitutes, assurance contracts or relative contribution incentives could be mechanisms that help solve collective action problems. Assurance contracts (à la Kickstarter) solve coordination problems, and only require payment should greater than some threshold X of agents commit to payment.⁸ Relative contribution incentives (Falkinger et al. 2002) reward agents based on how much more they contribute than the average within their income bracket. Butarin, Hitzig, and Weyl (2018) propose a quadratic contribution scheme that implements the efficient level of public goods in an incentive-compatible way. What auction theory did for online pricing, public goods mechanisms could do for the future of collective action. Unions or other organizations could take advantage of existing payments infrastructure (e.g., the union benefits debit card or points incentives) to implement these incentives. Union leaders with incentives to mobilize existing members around contract negotiation time could be partnered with to experiment with some of these mechanisms.

Increasing the returns to collective action

The other side of the equation is making the collective action that unions can generate more effective. Ultimately, this will require the restoration of an effective threat of raising employers' costs to intolerable levels and forcing bargaining over profits. Workers in a few key sectors still have this power, as in the public, health, and transportation sectors. But in an economy where the immediate employer's profit is passed up the value chain to other input suppliers like land, intellectual property, or capital, it means that organizations of workers might not have as their primary target the direct employer, but rather entities further upstream (as in the Justice for Janitors campaign, which went after the building that the janitors' employers serviced rather than the employers themselves). This would require

a large change in the legal architecture of collective bargaining, moving the level of bargaining away from employer-employee and encapsulating the whole value chain, including the financial entities at the top of it.

What collective action can accomplish also varies with economic fundamentals. The traditional view is that union leverage came via the strike; depriving the employer of labor was the costliest thing the organization of workers could do. Almost since the beginning of the modern labor movement, the strike has been hemmed in by the courts (Pope 2004). The 1937 *Fansteel* decision eliminated the sit-down strike as a tactic, asserting employer property rights over workers' right to strike. Since the 1938 *Mackaye* Supreme Court decision, employers can legally hire permanent replacement workers during a strike. These judicial decisions have been an effective check on the right to strike effectively, albeit with a lag (the use of permanent replacements accelerated greatly in the 1980s). Modern unions have adapted to this weakening of the strike as well as the other changes in the economy, and use both consumer pressure as well as capital market pressure (Webber 2018) in order to force employers to concede in contracts. A 21st century source of leverage might be control over data generated in the workplace; unions might find a role as stewards of data (e.g., value-added scores) generated by their members, and withhold access as a tool to secure higher pay. If automation is on the horizon, collective bargaining agreements can ensure that the productivity gains are shared with incumbent workers, blunting incentives for excess automation.

At the end of the day, reasonable people, facing the choice of whether to join a new union, will look down the decision tree and see if it passes a cost-benefit test. If the union cannot effectively pressure a company, they cannot win a good first contract; if no first contract, no wage premium; if no wage premium, no reason to risk signing an NLRB petition or voting against your employer. The ultimate determinant of union power is the capacity to use collective action to threaten firm profits; even density is subordinate to this basic capacity.

And here policy can potentially do a lot. One legal change is banning (i.e., making criminal) hiring replacement workers during strikes (which was rare prior to the 1980s); plenty of evidence finds that the ability to hire replacement workers during strikes is one of the largest contributors to strikes losing. Another way around

this is to eliminate exclusive representation and allow minority unionism, where a subset of workers can get legal recognition and strike protection without needing the whole firm. One implication of monopsony is that a good share of profits comes from the rents extracted from inframarginal workers; protected minority strikes would cut those profits and be even costlier than under competitive labor markets. A third way is to allow secondary strikes and boycotts; the complexity of the value chain makes solidarity across industries more valuable than before.

We might be surprised by a wave of militancy that sweeps American private sector workers back into union-like organizations. More likely, however, is that mass unionization can only come back alongside other large, difficult-to-anticipate political/economic transformations. In the interim, a need and demand for organizations that leverage the shared experiences of work and curb employer power remains, and policymakers and social scientists can help worker organizations meet that demand.

Suresh Naidu is an Associate Professor of Economics at Columbia University. Contact: sn2430@columbia.edu

Endnotes

- 1 Janus v. AFSCME is the 2018 Supreme Court case that eliminated mandatory union dues in public sector unions.
- 2 Jaumotte and Osorio-Buitron (2016) provide cross-country panel evidence from the OECD that union density is negatively correlated with inequality, despite widely varying industrial relations systems across the various countries. They instrument union density with the presence of the Ghent system of administering unemployment benefits via unions and lagged unemployment.
- 3 See Gourevitch (2014) for an exploration of the 19th century labor movement's criticisms of the labor market.
- 4 Farber et al. (2018) find a remarkably stable household union premium of 15%–20%, which is consistent with constant returns and non-unionized firms facing a residual labor supply elasticity of roughly 4, which is the upper bound estimated in the literature (Sokoleva and Sorensen 2018). Unlike the labor demand elasticity, there is no reason to think the supply elasticity facing the firm depends on union density or composition.
- 5 See Eidlin (2018) for an exploration of the differences between Canadian and US labor movements and political institutions.
- 6 In the US union membership closely tracks coverage, but this is different in other OECD countries (e.g., France) where density is low but coverage is high. As right-to-work laws expand in the US, the gap between coverage and membership may increase.
- 7
- 8 Dominant assurance contracts do even better, and make a transfer to agents in the event that fewer than X sign up to contribute, and can guarantee implementation of the efficient outcome in Nash strategies.

References

- Ash, Elliott, Bentley Macleod, and Suresh Naidu (2018) “The Language of Contract: Power and Promises in Collective Bargaining Agreements”
- Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. “Who’s who in networks. Wanted: The key player.” *Econometrica* 74.5 (2006): 1403-1417.
- Buterin, Vitalik, Zoe Hitzig, and E. Glen Weyl. “Liberal Radicalism: Formal Rules for a Society Neutral among Communities.” *arXiv preprint arXiv:1809.06421* (2018).
- Caldwell, Sydnee, and Nikolaj Harmon (2018) “Outside options, Bargaining, and Wages: Evidence from Coworker Networks”
- Caldwell, Sydnee, and Oren Danieli (2018) “Outside Options in the Labor Market”
- Cantoni, Davide, et al. “Are Protests Games of Strategic Complements or Substitutes? Experimental Evidence from Hong Kong’s Democracy Movement.” No. w23110. National Bureau of Economic Research, 2017.
- DiNardo, John, and David S. Lee. “Economic impacts of new unionization on private sector employers: 1984–2001.” *The Quarterly Journal of Economics* 119.4 (2004): 1383-1441.
- Eidlin, Barry. *Labor and the Class Idea in the United States and Canada*. Cambridge University Press, 2018.
- Falkinger, Josef, et al. “A simple mechanism for the efficient provision of public goods: Experimental evidence.” *American Economic Review* 90.1 (2000): 247-264.
- Farber, Henry S., et al. “Unions and Inequality Over the Twentieth Century: New Evidence from Survey Data.” No. w24587. National Bureau of Economic Research, 2018.
- Frandsen, Brigham. “The surprising impacts of unionization on establishments: Accounting for selection in close union representation elections.” Manuscript. Department of Economics, Brigham Young University (2013).
- Gerber, Alan S., and Donald P. Green. “Field experiments on voter mobilization: An overview of a burgeoning literature.” *Handbook of Economic Field Experiments*. Vol. 1. North-Holland, 2017. 395-438.
- González, Felipe. “Collective Action in Networks: Evidence from the Chilean Student Movement.” (2017).
- Gourevitch, Alex. *From slavery to the cooperative commonwealth: labor and republican liberty in the nineteenth century*. Cambridge University Press, 2014.
- Hensel, Lukas, et al. “Political Activists as Free-Riders: Evidence from a Natural Field Experiment.” (2018).
- Jacome, Elisa, Ilyana Kuziemko and Suresh Naidu (2018) “Intergenerational Mobility over the 20th Century: Evidence from Small Data”.
- Jaumotte, M. F., & Osorio, M. C. (2015). *Inequality and labor market institutions*. International Monetary Fund.
- Lee, David S., and Alexandre Mas. “Long-run impacts of unions on firms: New evidence from financial markets, 1961–1999.” *The Quarterly Journal of Economics* 127.1 (2012): 333-378.
- Naidu, S., & Reich, A. (2017). Collective Action and Customer Service in Retail. *ILR Review*.
- Pope, James Gray. “How American workers lost the right to strike, and other tales.” *Mich. L. Rev.* 103 (2004): 518.

- Schmitt, John, and Ben Zipperer. "Dropping the ax: Illegal firings during union election campaigns." Washington, DC: Center for Economic and Policy Research, 2007.
- Sokolova, Anna, and Todd Sorensen. "Monopsony in Labor Markets: A Meta-Analysis." (2018).
- Webber, David. *The Rise of the Working-Class Shareholder: Labor's Last Best Weapon*. Harvard University Press, 2018.
- Weil, David. *The Fissured Workplace*. Harvard University Press, 2014.

How to think about finance?

Atif Mian*

There has been a major structural shift in financial markets since the 1980s. The world is awash in credit, and credit is cheaper than ever before. I discuss how increasing financial surpluses within parts of the economy have resulted in an expansion in the supply of credit, which has largely financed the demand-side of the real economy. This increasing reliance on “credit as demand” raises some serious policy questions going forward. I discuss the importance of equitable and inclusive growth, fair taxation system and risk-sharing in creating a financial system that promotes prosperity and stability.

The big shift in finance

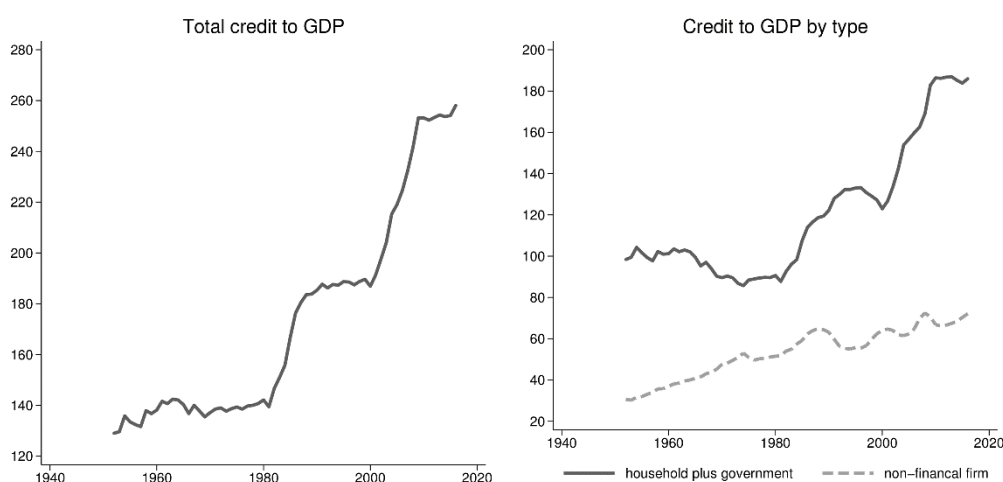
The left panel of figure 1 shows that total credit in the U.S. remained relatively flat at around 140% of GDP in the post-war years until 1980. Since 1980, however,

credit has gone up at a rapid pace, reaching a historic high of 258% of GDP in the most recent numbers available for 2016¹. Even the great recession did not put much of a dent in the growth of credit, with total credit rising from 232% in 2007 to 258% in 2016.

The phenomenal rise in credit is in fact a global phenomenon as the work of Oscar Jorda, Moritz Schularick and Alan Taylor has carefully documented. The recently released global debt database from the IMF that covers both advanced and emerging economies also shows a big increase in global credit to GDP from around 150% between 1960 and 1980 to over 250% in 2016.

The right panel of figure 1 splits total credit into non-financial firm credit and credit going to households plus government. The figure shows that most of the increase in credit since 1980 has been driven by credit going to households and the government. Credit going

Figure 1 Big shift in finance



to the corporate sector plays a relatively minor role in explained the big rise in total credit. In particular, 82% of the increase in total credit as a share of GDP since 1980 is driven by credit going to households and the government. The big increase in total credit, especially credit going to households and government, is not unique to the U.S. Jordà et al. (2016) show that there has been a large increase in private bank credit in all advanced countries since 1980. The authors further show that the increase in credit is dominated by mortgage credit going to the household sector.

The large increase in credit has been accompanied by a persistent decline in the price of credit as long term interest rates have fallen to historic lows. For example, the average 10-year real interest rate has declined from a high of 6% in 1983 to zero in recent years (IMF WEO (2014)). The fall in the price of credit even as the quantity of credit exploded suggests that the expansion in financial sector is driven by an increase in the “supply” of credit. What is behind this structural shift in finance?

The United States, and the global economy, experienced a couple of major structural shifts around 1980. First, share of income going to the top 1% of earners went up significantly. The richest 1% of Americans captured 11% of total income in 1980 and 20% in 2014 (Piketty and Saez (2003)). The rise in top income share occurred in many other countries as well (see e.g. Alvaredo et al. (2018)). Second, the collapse of the Bretton Woods system in 1971 ushered in the era of global capital flows. A number of countries started running large current account deficits or surpluses post 1980. High savings by oil-rich countries and high-growth Asian economies have also contributed to the global rise in credit through cross-border flows that have accelerated since 1980. The

U.S. had close to a balanced current account in 1980 but has been consistently running current account deficits of over 2% of GDP since then.

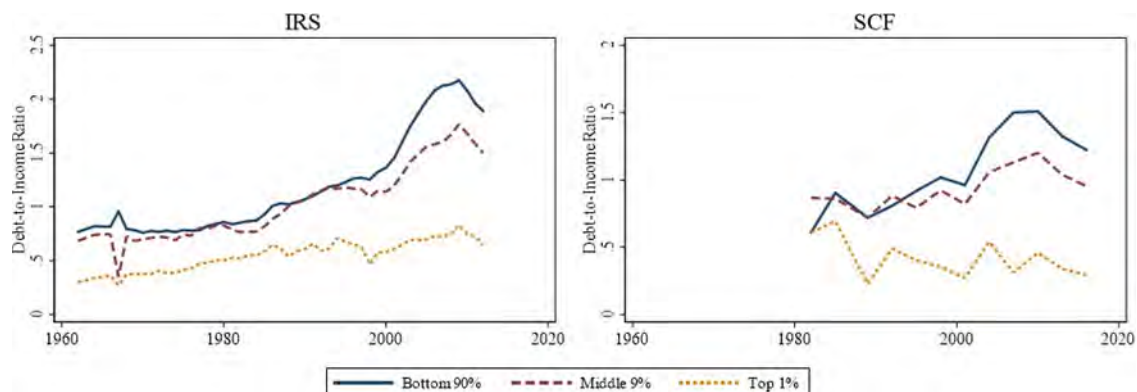
The rise in top income share contributes to the expansion of the size of the financial sector. High-income earners save a large share of their income, creating a larger “financial surplus” within the economy that is then channeled back through the financial sector. The financial sector deploys these larger gross savings back into the economy through credit creation. There is a close association between the rise in top-income share of the top 1% and the rise in household leverage for the rest of the population, suggesting that increased gross savings from the top 1% were partly absorbed by increased borrowing by the remaining household sector.

Figure 2 shows the evolution of debt to income for the top 1%, middle 9% and the remaining bottom 90% in both IRS and SCF data sets. The rise in household credit is concentrated in the bottom 99% and not the top 1%, while income gains since 1980’s have largely gone to the top 1%. Mian and Sufi (2018) show the same pattern holds when using individual level credit bureau data.

What has the increased credit supply financed? The “credit as demand” channel

The increasing financial surpluses, or savings gluts, have expanded the total supply of credit to the economy, lowering long-term interest rate in the process. What has the increased supply of credit financed?

Figure 2 Debt-to-income ratio across the income distribution



The textbook model of finance says that credit is used to finance real investment: savers deposit their surplus funds in the banking sector which then lends these funds to firms for investment. In other words, credit is used to finance production, or the supply-side of the economy. However, evidence suggests that a relatively small fraction of the increase in credit has gone towards funding production. For example, despite the large increase in credit creation, rate of investment has not gone up. The average U.S. gross investment rate was 22.5% from 1947 through 1979 and 21.8% from 1980 onwards.

Other evidence is also at odds with the idea that the additional credit has gone into increasing productive capital. Overall growth is not any higher post-1980. Moreover, there is strong evidence that productivity growth has slowed down significantly over the last decade and a half. If additional credit has not gone into financing production as much, then the other possibility is that credit has increasingly been used to fund demand. There is indeed robust evidence to support this view.

I have already shown that most of the increase in credit since 1980 has been used to fund government fiscal deficits, or household financial deficits, especially households outside of the top 1%. The concentrated growth in government and household debt suggests that aggregate demand is increasingly reliant on credit creation for support.

The reliance on credit creation for supporting aggregate demand is a natural consequence of a higher share of income being saved due to increased inequality. Equilibrium condition for the real economy implies that as a larger fraction of the output is saved, the increased savings must be channeled back to the real economy either as investment or consumer demand. In the absence of such a channel, the real economy will be forced to contract - or not grow as fast - in order to equate supply and demand in the real economy. This phenomena is sometimes referred to as “liquidity trap” or “savings trap” in the literature (e.g. Eggertsson and Krugman (2012)).

Theoretically, as long as certain sectors within the economy such as the government or households below the top 1% are willing to run larger deficits, the real economy can continue to grow at full capacity. However, as the economy continues to rely on credit-creation for supporting demand, it becomes increasingly more

difficult to do so. The reason is that as household and government credit builds up, interest rate needs to fall in order to keep the debt service requirement manageable. The reduction in interest rate also tends to raise asset prices, especially housing values, which enables household to borrow more easily. But the dependence on ever lower interest rate to support a larger stock of debt cannot go on forever.

At some point it becomes difficult for interest rate to drop any further without adding a cost of its own. First, there is the natural zero lower bound constraint on nominal interest rate. Second, and perhaps more importantly, very low interest rates introduce other problems that are damaging for the overall economy. For example, asset markets are more prone to bubbles at very low interest rate. It becomes increasingly difficult to fund pension plans and insurance funds with long-dated liabilities.

The combination of high debt and increased likelihood of bubbles makes financial sector more fragile. Low interest rates can also inhibit productivity growth due to greater misallocation of capital (Gopinath et al. (2016)) or increased market concentration (Liu et al. (2018)).

Long-run policy implications

The remarkable growth in credit and the accompanying fall in long-term interest rate since 1980 represents the most important shift in finance in the modern era. The discussion above highlights why this shift is not sustainable, at least not without major harm to economic growth. A reliance on continuous credit creation to generate demand eventually slows down economic growth through liquidity trap like scenarios and other ill effects of very low interest rates.

What can be done to reduce the dependence on credit and create more space for economic growth as a result? As I mentioned, the root causes of secular credit growth lie in large financial surpluses in the economy that are then channeled through the financial system. A reversal of excessive credit dependence requires that financial surpluses be brought down to healthier levels. There are three types of structural changes in the economy that can help reduce the dependence on credit creation for aggregate demand.

First, more equitable growth will reduce the excessive savings that are accumulated by the top 1%. As explained above, there is a direct relationship between highly skewed economic growth and a bloated financial sector that results in broader economic malaise. Second, estate taxes and wealth tax (e.g. as proposed by Piketty (2014)), especially on “money-like” instruments, can be useful in restraining excessive surpluses. Some of the revenue raised from wealth and estate taxes can be used to lower income taxes for lower income brackets that have a high propensity to spend. Third, high inter-generational mobility helps to reduce the adverse effects of financial surpluses as accumulated surpluses naturally get liquidated across generations. Thus policies that strengthen public education and provide more equitable opportunities to the entire population help reduce dependence on credit creation.

Cyclical and more immediate policy implications

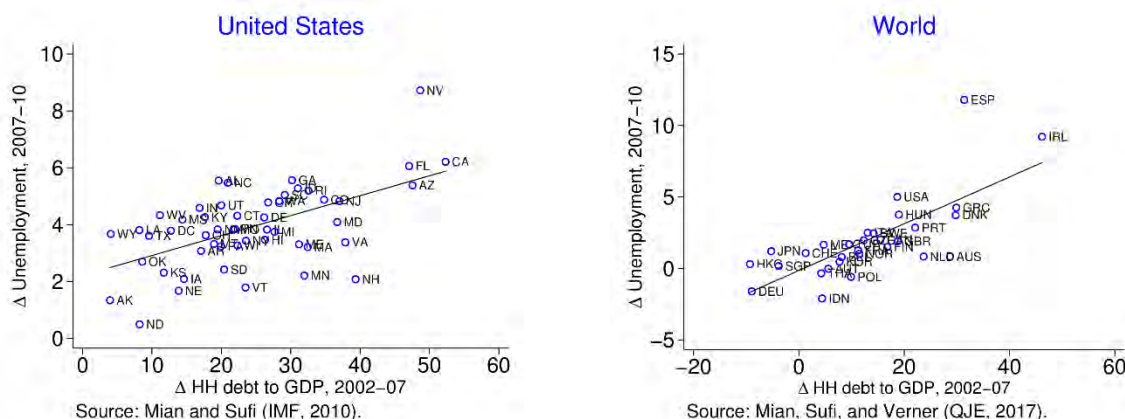
I next turn my attention to more immediate steps that can be taken to reduce the cyclical costs of problems emanating from financial markets. The most striking empirical regularity connecting credit and business cycles in recent decades is that large run up in credit, especially household credit, tends to be followed by an increase in unemployment. The 2008 global crisis, as shown in figure 3, was one manifestation of this broader trend. States within the U.S. that had a larger increase in household leverage between 2002 and 2007, ended up experiencing a much more severe recession. Remarkably, we find exactly the same relationship across countries.

In Mian and Sufi (Forthcoming), we explain how recurrent business cycle contractions such as the one in figure 3, are the result of “credit-driven household demand channel”. The basic idea is that expansion in the supply of credit fuels a boom in credit and asset prices that boosts household aggregate demand. However, the expanding credit boom also sows the seeds of its own destruction and ultimately results in a macroeconomic slowdown.

How should policy be tailored to address such credit-induced boom-bust cycles? I discuss steps regarding macro-prudential policy, tax policy, banking regulation, GSE reforms, and bankruptcy law that help reduce the likelihood and adverse consequences of credit-induced boom-bust cycles. On the macro-prudential front, the most important policy focus should be to facilitate better risk-sharing between creditors and borrowers. Credit creates problems for the macro-economy in the event of a downturn due to differences across creditors and debtors in their marginal propensity to spend. A downturn naturally hurts borrowers disproportionately more since they are levered. Borrowers are also more sensitive to shocks as they tend to have much higher propensity to respond to shocks. The combination of these two forces implies that for any given macro shock, headwinds faced by the macro economy are stronger the more levered the economy is. Moreover, households may not fully internalize the possibility of such headwinds when deciding how much leverage to undertake. This results in economies getting “over-leveraged” with deeper and more frequent recessions.

A natural solution to minimizing the disruptions caused by credit is to promote ‘state-contingent contracting’ that allows a more equitable sharing of downside risk

Figure 3 Credit growth and recessions



between creditor and debtor in the event of a macro downturn. The creditor will naturally be compensated for sharing downside risk upfront. State-contingent contracting does not suffer from the usual moral hazard problem in risk-sharing contracts, because the risk-sharing is contingent upon a macro state of nature over which the borrower has no direct control.

There are multiple examples of state-contingent contracts that have been proposed in the past. For example, sovereign debt repayment can be linked to GDP as proposed by Robert Shiller. Student debt repayment can be linked to the earning potential of a graduating student's cohort and major. We proposed a state-contingent "shared responsibility mortgage" (SRM) in our book Mian and Sufi (2014). SRM works by reducing monthly mortgage payments in the event of a local downturn in housing market without altering the amortization schedule - effectively providing both cash-flow and principal relief for borrowers.

The promotion of state-contingent contracts will have multiple advantages at the macro level. First, it will significantly reduce real macroeconomic volatility through the introduction of automatic stabilizers as debtors and creditors share risk more efficiently. Economic volatility is especially harmful for lower income households with more fragile economic conditions. Second, it will raise total welfare by avoiding long period of times when economy operates below capacity due to the after effects of debt overhang. In this way an economy with state-contingent contracting is both more resilient and stronger. Empirical evidence coming out of the Great Recession shows that better risk-sharing between debtors and creditors would have significantly reduced the extent and scale of the recession. Di Maggio et al. (2017) show that lower interest rates post-2008 were not passed-through to many constrained households who were unable to refinance, thus putting a real drag on aggregate demand. Ganong and Noel (2017) show that reduction in mortgage payments under a government program significantly increased spending and lowered defaults. More than four million homes were foreclosed over a short period of time during the Great Recession. Mian et al. (2015) show these fire sales put further downward pressure on house prices, thus worsening an already bad situation. State-contingent mortgages would have helped reduce the number of homes going into foreclosure.

A natural question that arises is that if state-contingent

contracts are so beneficial, why do we not see more of them around us? There are three main reasons for this. First, the argument in favor of state-contingent contracting is based on the negative macro externalities inherent in standard debt contracts. Private agents, for both rational and behavioral reasons, are not likely to internalize these externalities. There is thus a rationale for promoting state-contingent contracting as part of macro-prudential policies. Second, as Admati et al. (2018) point out, shareholders have an incentive to ratchet leverage up since some of the benefits of reducing leverage accrue to creditors and not shareholders. Third, there are a number of institutional features in the U.S. and abroad that hinder the adoption of state-contingent contracting. I discuss specific policy steps that remove such obstacles and encourage adoption of state-contingent contracts.

The U.S. tax system offers an interest expense deduction that reduces the effective cost of debt financing for homeowners. The tax subsidy is naturally capitalized in people's housing decisions and the value of the housing market. The tax subsidy distorts the financial system by encouraging leverage that is harmful from a macro-prudential perspective. Removing the tax subsidy is not feasible politically, and doing so may also depress the housing market. Therefore, we proposed in Mian and Sufi (2014) that the subsidy be moved over towards state-contingent contracts like the SRM that have nice macro-prudential characteristics. The current system not only subsidizes the housing sector, but does so in a way that is harmful from a prudential perspective.

Bank capitalization rules under the Basel system are also structured to discourage banks from holding state-contingent securities in their portfolio. For example, suppose a bank issues a traditional mortgage of 100,000\$ with 80% loan to value ratio. How much capital does the bank need to issue this mortgage? The typical capitalization requirement is 8%, implying the bank needs 8,000\$ of capital. However, Basel rules would give this mortgage a "risk weight" of 0.3, meaning that the bank only needs to cover 0.3 times the 8%, or 2,400\$ in capital.

Now imagine the bank issued the same mortgage as an SRM. The risk weight would increase substantially, probably close to 1. As a result, the bank would need to have 8,000\$ in capital, instead of the 2,400\$, to issue the mortgage as an SRM. The higher capital requirement for state-contingent contracts is unfortunate and

somewhat ironic from a societal view point.

The banking system is designed to discourage banks from holding state-contingent contracts that are more beneficial from a macro perspective. The premise behind banking regulations such as Basel III is that losses must be minimized for the banking sector, or creditors at large. The banking system is therefore encouraged to originate the “safest” of assets from the creditors’ perspective, and pass on all risk of debtors. However, as I have already explained, doing so leads to much worse outcomes in the event of a cyclical downturn. The current structure of banking regulation does not split risk between creditors and debtors in a socially beneficial manner.

The adoption of a new “standard” in financial markets often requires the government to step in and define new rules. For example, there is now an active and liquid market for inflation-indexed treasuries or TIPS. But that market was created by the U.S. government itself under Clinton administration. Similarly, the 30-year fixed rate mortgage became the standard mortgage in the U.S. after the government actively encouraged it. Today government-sponsored entities (GSEs) are by far the largest players in the mortgage origination business. The explicit government support enjoyed by ‘conforming’ mortgages supported by the GSEs means that it is more difficult for the private sector to introduce new solutions like state-contingent mortgages.

Given the existing large role of government in mortgage origination and the societal benefits associated with state-contingent contracts such as SRMs, the government could include state-contingent mortgages in its definition of ‘conforming’. The government could also help in defining states of the world, such as official local house price indices, to promote state-contingent mortgages. State-contingent contracting is an example of ex-ante macro prudential intervention. If properly implemented, it has the virtue of endogenously reducing economic volatility and crises, and hence the need for ex-post intervention in the first place. However, to the extent ex-post intervention is needed, efficient bankruptcy laws help in dealing with debt-overhang. The U.S. has better bankruptcy laws compared to rest of the advanced world, especially for households. However, there are certain areas, such as student loans, where bankruptcy laws need to be amended to enable restructuring of odious student debt.

The broader message to take-away is that risk-sharing between creditors and debtors is an important principle to promote. Macro-prudential and regulatory policies should be designed to favor risk-sharing. Unfortunately the current tax and regulatory system is designed to do the opposite. The present regulatory regime is “bank-centric”, with an exclusive focus on minimizing default probability for banks. This is short-sighted and does not take into account the true cost for the real economy when the financial system passes risk squarely on debtors. The banking sector should be better capitalized, as argued forcefully by Admati and Hellwig (2013), with a capital structure that is more suited to absorbing losses without going into bankruptcy.

Atif Mian is the John H. Laporte, Jr. Class of 1967 Professor of Economics, Public Policy and Finance at Princeton University. Contact: atif@princeton.edu

Endnotes

* This essay was prepared as part of the Economics for Inclusive Prosperity (EfIP) series of policy briefs. I thank Anat Admati and Dani Rodrik for helpful comments.

1 Total credit includes private credit (household debt and non-financial firm debt) and sovereign credit.

2 This would be akin to having a negative interest rate on some margins in the current environment.

3 For broader evidence on household demand channel constraining the efficacy of monetary policy, see Agarwal et al. (2017, 2018); Aladangady (2014); Baker (2018); Cloyne et al. (2017); Jordà et al. (2014)

References

- Admati, Anat and Martin Hellwig, *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It*, 1 ed., Princeton University Press, 2013.
- Admati, Anat R., Peter M. Demarzo, Martin F. Hellwig, and Paul Pfleiderer, "The Leverage Ratchet Effect," *The Journal of Finance*, 2018, 73 (1), 145–198.
- Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru, "Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program," *Journal of Political Economy*, 2017, 125 (3).
- , Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel, "Do Banks Pass Through Credit Expansions to Consumers Who Want to Borrow? Evidence from Credit Cards," *Quarterly Journal of Economics*, 2018, 133 (1), 129–90.
- Aladangady, Aditya, "Homeowner Balance Sheets and Monetary Policy," FEDS Working Paper, Federal Reserve Board 2014.
- Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, "World Inequality Report," 2018. Working Paper.
- Baker, Scott R, "Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data," *Journal of Political Economy*, 2018.
- Cloyne, James, Clodomiro Ferreira, and Paolo Surico, "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism," 2017.
- Di Maggio, Marco, Amir Kermani, Benjamin J. Keys, Tomasz Piskorski, Ramcharan Rodney, Amit Seru, and Vincent Yao, "Interest Rate Pass-Through: Mortgage Rates, Household Consumption, and Voluntary Deleveraging," *American Economic Review*, 2017, 107 (11), 3550–88.
- Eggertsson, Gauti B. and Paul Krugman, "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach," *Quarterly Journal of Economics*, 2012, 127 (3), 1469–1513.
- Ganong, Peter and Pascal Noel, "Consumer Spending During Unemployment: Positive and Normative Implications," 2017. Working Paper.
- Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez, "Capital Allocation and Productivity in South Europe," Working Paper 21453, National Bureau of Economic Research August 2016.
- Jordà, O`scar, Moritz Schularick, and Alan M. Taylor, "Betting the House," 2014. Federal Reserve Bank of San Francisco Working Paper.
- Jordà, Òscar, Moritz Schularick, and Alan M. Taylor, "The great mortgaging: housing finance, crises and business cycles," *Economic Policy*, 2016, 31 (85), 107–152.
- Liu, Ernest, Atif Mian, and Amir Sufi, "Low Interest Rates, Market Power and Productivity Growth," 2018.
- Mian, Atif, Amir Sufi, and Francesco Trebbi, "Foreclosures, house prices, and the real economy," *The Journal of Finance*, 2015, 70 (6), 2587–2634.
- and —, *House of Debt: How They (and You) Caused the Great Recession and How We Can Prevent It From Happening Again*, University of Chicago Press: Chicago, 2014.
- and —, "Inequality, Credit Creation And Aggregate Demand," 2018. Working Paper.
- and —, "Finance and Business Cycles: The Credit-Driven Household Demand Channel," *Journal of Economic Perspectives*, Forthcoming.
- Piketty, Thomas, *Capital in the Twenty First Century*, 1 ed., Harvard University Press, 2014.
- and Emmanuel Saez, "Income Inequality in the United States, 1913–1998*," *The Quarterly Journal of Economics*, 2003, 118 (1), 1–41.

Labor in the Age of Automation and Artificial Intelligence

Anton Korinek*

As technology advanced in recent decades, it increasingly left workers behind and led to sharp increases in inequality. The current wave of progress in artificial intelligence is likely to accelerate these trends. This note lays out three complementary approaches to countering these developments. Firstly, since technological progress generates net gains for society as a whole, the winners could in principle compensate the losers and still be better off. Secondly, progress should be steered to minimize the losses of workers. Thirdly, there is an important role for government intervention in information technology to thwart the rise of monopolies that extract rents from society. The note concludes with some speculations on the impact of artificial intelligence increasingly rivaling human labor.

Introduction

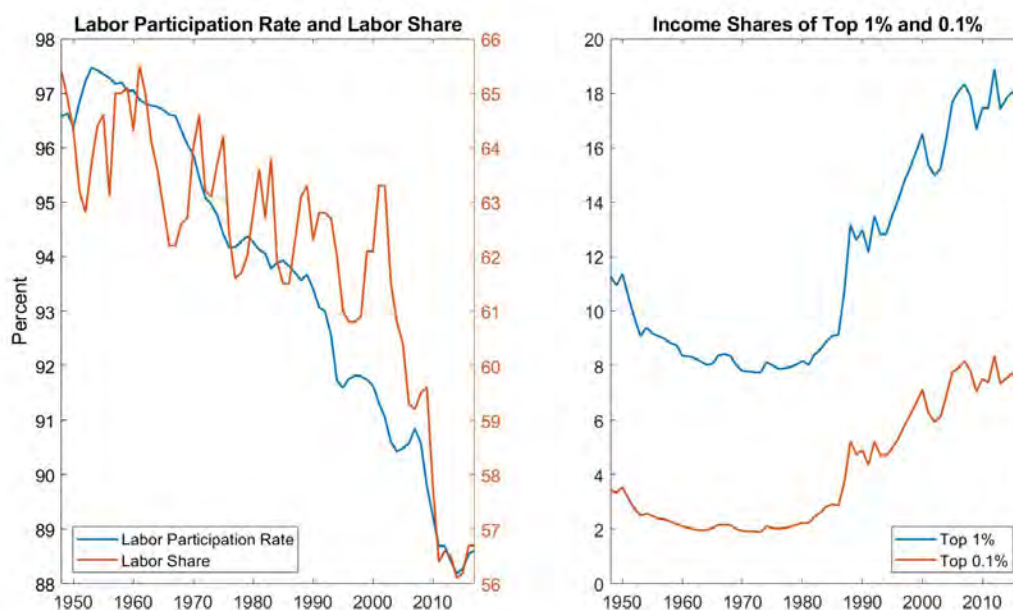
By and large, workers have had a good run over the first two centuries of the Industrial Revolution. Technological progress automated tasks involving hard physical labor and tedious routine activities and, as a by-product, increased workers' incomes about ten-fold. In the eyes of idealists, progress freed human workers from tasks that were in fact inhumane, that humans were never meant to perform – so they can focus on more fulfilling work involving cognitively interesting activities.

Starting about four decades ago, however, technological progress has increasingly left workers behind, as reflected in a range of dismal statistics: Since World War II, the labor force participation rate of prime-age men has declined from 98% to 89%; and the labor share

of income – what is earned by workers rather than capitalists – has declined from 66% to 58% (see left panel of the figure 1). The average real wage of regular workers has in fact declined over the past four decades – over a period in which total income in the US almost tripled! At the same time as regular workers fell behind, the so-called superstars of the economy have garnered an increasing share of income, with the top 1% more than doubling their take to about 20% of all income, and the top 0.1% tripling their take to close to 10% (see right panel of the figure 1). Looking at wealth rather than income, some estimates suggest that the richest three Americans now own more than what the bottom 50% of the US population own.

Although technological forces were undeniably a prime force behind these developments, there were also other factors involved, many of which are discussed

Figure 1



comprehensively by other policy briefs of this series by Economists for Inclusive Prosperity. For example, trade liberalization put pressure on workers competing with cheaper labor abroad. Institutional changes such as less generous redistributive policies, the declining power of unions, and tax policies favoring the rich reduced the take-home income of regular workers. Many of these factors were in fact also partly driven by technological forces.

This brief focuses squarely on the implications of technological change and how to manage them. The traditional approach of economists has been to view technology as a driving force that is outside the realm of our analysis – technology is developed by engineers; we economists take it as given and study the implications for the economy. But technology is not destiny. In fact, better understanding the technological forces behind the decline of labor is crucial for shaping our agenda on how to best protect workers going forward. In the following, I will discuss the broader forces that have contributed to rising inequality in recent decades and how to counter-act them; I will zoom in on the implications of information goods and digital technologies for the economy; finally, I will speculate on how the rise of artificial intelligence will affect workers and the economy in coming decades. In each section, I will include a discussion of the policy options available.

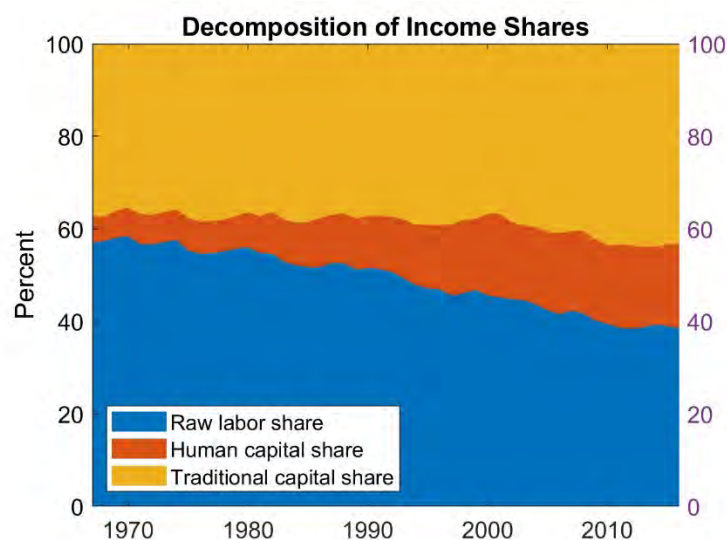
Does Technological Progress Replace Workers?

It has been true since the advent of the Industrial Revolution that technological progress replaced specific jobs – at the time, for example, spinners and weavers. But time and time again, after a period of adjustment, the economy created new jobs for the displaced workers that ultimately paid better. (If the disruption was severe, the adjustment took longer, but at least the children of displaced workers found better jobs.) Many economists therefore proclaim that technological progress is unambiguously good for workers.

However, it is not a natural law that technological progress will lead to higher wages and improved livelihoods. The fact that real wages of regular workers have declined over the past four decades, strongly suggests that the overall effect of technological *change* over that period (one hesitates to call it technological progress) has been to reduce the market wages of regular (unskilled) workers. As technology advanced in recent decades, the economy simply seemed to have less and less need for unskilled labor.

At first sight, the picture looks better for skilled workers who saw their wages rise significantly over the 1980s and

Figure 2



1990s, although less so in the recent two decades (see e.g. Autor, 2015; Brinca et al., 2019). However, there is also a bleaker interpretation of this phenomenon: skilled labor can be interpreted as unskilled labor enhanced by education, i.e. it is a composite of unskilled labor and human capital. The wages of skilled workers can, in this view, be decomposed into the wages of unskilled workers plus returns on human capital. Since the wages of the unskilled have not increased, all of the increase in skilled wages in fact reflects returns on human capital investment.

Figure 2 above decomposes what fraction of national income was earned by unskilled labor versus capital (made up of both traditional and human capital) from 1967 to 2017. The “human capital share,” calculated as the extra returns earned by college graduates, has risen from 5.6% to 18.2% of total income. Conversely, the raw labor share has declined from 57% to less than 40%. According to this interpretation, combined capital earns more than three fifths of all output in the economy.

From this point of view, the past four decades have led to an even starker reallocation of returns from labor to capital (which now includes human capital). The difference matters for workers, since wage earnings reflect the return on human labor effort, whereas the returns on human capital are returns on investments in education, which is becoming ever more costly and which in fact soaks up a large part of these returns.

Technological Redistribution and Social Redistribution

When technological progress leads to income losses for workers, it is natural to think about ways to compensate them for their losses. An important point to emphasize is that technological progress could in principle make everybody better off, i.e. generate what economists call a Pareto improvement, if there is sufficient political will. By definition, technological progress means that the economy can produce more for a given amount of inputs, implying that there is more overall income to be distributed. If one of the factors of production, for example unskilled labor, earns less as a result of an innovation, it means that someone else is not only earning the additional fruits of progress, but also appropriating part of what used to be the earnings of unskilled labor.

More generally, we can decompose the economic effects of technological progress into two parts, as laid out e.g. in Korinek and Stiglitz (2019): First, progress raises overall output, i.e. it increases the size of the economic pie produced. This extra output is earned by someone in the economy, for example by the innovator who reaps the fruits of her innovation. Secondly, technological progress also generates a redistribution of the existing economic pie, as it changes the market prices at which people transact in the economy. For example, it may reduce the wages of some workers and increase the wages

of other workers. This redistribution via price changes is always zero-sum: price increases benefit sellers at the expense of buyers, and vice versa for price reductions. We may call this effect a technological redistribution of income since it is generated by technological forces playing out in the market economy.

Consider for example a new AI system that replaces human radiologists: the first effect of such a system may be to lead to better diagnosis and increased use of radiology services, increasing the size of the economic pie produced. The second effect may be to reduce the market wages of human radiologists who had specialized in interpreting images but to increase the wages of nurses and of specialists who rely on radiologists, who can now do more without requiring input from costly human radiologists.

In an idealized world, if we want to avoid that technological progress leaves behind some members of society, then social redistribution would aim to undo the described technological redistribution to compensate the losers of innovation while keeping the increased size of the pie. Those who benefit from technological redistribution accrue windfall gains, i.e. gains that are not based on their own efforts but more on luck. From a policy perspective, this makes it important to be explicit about who are the beneficiaries of technological redistributions, and to look for ways to tax their windfall gains to compensate the losers. At times, it may be possible to tax away such windfall gains without introducing the distortions that taxes usually generate. In those cases, undoing technological redistribution may be feasible without any efficiency losses to the economy. For example, if an innovation increases the value of land in particular areas, higher property taxes could tax away these windfall gains. Korinek and Stiglitz (2018) show that this is a more efficient solution than e.g. the “robot taxes” that have been proposed.

However, there are also many technological redistributions that are quite difficult to undo in practice. In our radiology example above, we would need to tax away the wage gains of nurses and other specialists to compensate radiologists for their losses. This is a proposition that is impractical because of a variety of information problems – it would require a system of taxes and transfers that is far more fine-tuned than what is possible, and it would introduce a number of moral hazard problems, for example reducing work incentives for the beneficiaries of technological

redistribution who would face higher tax burdens. Furthermore, some may also view it as unfair if nurses are taxed to compensate better-earning radiologists for the losses stemming from technological redistribution. In all those cases, a general progressive tax system – one that charges higher rates to individuals earning more – together with a social safety net that limits the downside for the losers of technological progress may be one of the best available second-best solutions.

Steering technological progress

Technological progress is the result of conscious and targeted efforts of innovators – unlike the way it is described in many of our economic models, in which progress just happens exogenously. When an innovator comes up with ingenious new methods of producing novel goods or services, or with novel processes to produce existing goods or services more cheaply, her incentives are set by market forces – but the price signals sent by the market do not always reflect social value. This phenomenon is well-known when it comes to externalities such as pollution, and there is wide consensus among economists to regulate such externalities and correct the price signals sent by the market to better reflect our social values.

If we as a society care about inequality and we cannot realistically achieve the desired income distribution purely via transfers, then it is natural to extend this framework of correcting price signals to make innovators more conscious of the distributive externalities of their inventions. For example, if an innovator comes up with a clever new technology to replace the work of thousands of unskilled workers with a handful of skilled workers, the innovation will create a large technological redistribution – unskilled worker will see their wages decline and skilled workers will experience wage gains, exacerbating the trends of the past four decades. Neither of the two groups of workers have actively contributed to these windfall gains and losses, so they constitute externalities. Economists have traditionally been skeptical of such arguments because the described externalities are so-called “pecuniary” externalities – they occur because market prices and wages adjust. If we only care about efficiency not equity, then it is desirable to ignore pecuniary externalities. If we also care about equity, then pecuniary externalities are at the center stage and need to be addressed to achieve our goals.

Technology policy should thus steer technological progress so as to encourage innovation that has desirable distributive properties and to discourage innovations that increase inequality. Let me outline three different avenues for doing this:

The first avenue is to focus on the distributive implications of all the research that is conducted or sponsored by government. Government is one of the largest sources of research funds in our economy, and it should actively steer progress in directions that augment workers rather than replacing them. One example of this is what has come to be called *intelligence assistance*, i.e. AI systems that are designed to complement and enhance the abilities of workers so they can perform higher value-added tasks. Such intelligence assistance may make it possible for workers to do jobs that were previously out of reach for them, greatly increasing the demand for unskilled labor. If intelligence assistance systems were privately funded, there is significant risk that their creators will reap most of the economic returns; if they are publicly funded, by contrast, they can be made available for free or at cost, and workers can reap the resulting returns.

The second avenue is to use regulatory powers as well as tax and subsidy schemes to steer technological progress, in a similar fashion to how other types of externalities such as pollution are addressed. If it is possible to identify whether a specific type of innovation will have positive or negative distributive effects, then the innovative activity itself could be subsidized or taxed, or patent lives on the respective innovations could be lengthened or shortened. Otherwise, subsidizing the employment of lower-skilled workers would lower the cost of such employment and provide socially more desirable price signals to innovators (just like putting a price on carbon induces innovation to engage in carbon-saving activities). For example, if unskilled labor becomes cheaper, then it is less desirable to develop innovations that save on unskilled labor.

A third avenue to steering the path of technological progress is simply to create more awareness of the distributive implications of different types of innovation. Although it is difficult to predict what the exact impact of an innovation on labor markets will be, there are some general guidelines: for example, process innovations that reduce costs by automating labor are more likely to hurt workers than product innovations that generate new products that meet previously

unknown needs. Many entrepreneurs are socially-minded and care about the impact of their innovations on society. Making them more aware of the distributive implications of their actions will make a difference. There is also a vibrant NGO sector in the US, partly funded by high tech billionaires, that could make it one of its priorities to invest in innovations such as intelligence assistance that complement unskilled workers rather than displacing them.

Digitization, information goods, and the rise of superstars

An aspect of the recent wave of technological progress that sets it apart from earlier waves of progress is that it centers on digitization and information goods. This is most visible in the IT sector, where some companies generate billions of dollars of revenue selling digital goods while employing just a few hundred employees (who are usually highly skilled) to produce them. However, information goods are a broad phenomenon that is increasingly relevant not only in the technology sector but throughout the economy: in sectors from retail to the food and beverage industry, productive companies such as Walmart or Starbucks replicate their success in local market after market by copying an information good – best business practices – over and over again.

What makes information different from tangible goods is that it is *non-rival* but *excludable*. Non-rival means that it can be used without being used up: if somebody writes a computer program, billions of people can use the same code without using it up. By contrast, tangible goods are typically rival and are eventually used up when they are used: if someone eats a loaf of bread, no one else can eat it. The non-rival nature of information goods implies that once a company has incurred the cost of developing them, it can copy them many times at negligible marginal cost. This means that sectors in which information goods play an important role are *natural monopolies*: it is most economical to develop an information good only once or (if tastes differ) a small number of times, and then to distribute it to the entire market.

If an information good is created by a private owner, the excludable nature of such goods implies that its owner can prevent competitors from using it and therefore

has market power. This enables the owner to charge higher prices and extract monopoly rents. In our paper on “The Macroeconomics of Superstars” (with Ding Xuan Ng, 2018), we argue that most of the rise of market power in recent decades and the associated rents can be explained by digitization and information goods across the US economy. This has also been an important factor behind the rise in inequality over the period.

Public policy faces two fundamental problems when dealing with information goods:

The first fundamental problem is that the private market has difficulty achieving efficient outcomes when information goods are involved. On the one hand, financing information goods necessitates that private companies have some monopoly power so they can charge a markup over their marginal cost and earn rents to recoup the cost of their investment. Our society typically grants such monopoly power by awarding intellectual property rights to the creators of information goods that provide them with exclusivity. On the other hand, the monopoly markups that such firms are charging imply that consumers face higher prices and will demand less than what is efficient. As a result, the private market will both underprovide and underuse information goods. Furthermore, it turns those who successfully commercialize information goods into so-called superstar firms, leading to large increases in inequality.

The most efficient solution in the face of these problems would be to publicly fund the creation of information goods and then – since they are almost free to copy – distribute them at a very low price (technically, at marginal cost) to anybody who is interested in using them. This model works relatively well for fundamental research. An example is when DARPA used public funds to finance the invention of the Internet, which has since created trillions of dollars of value. The role of government in financing information goods and making them freely available to society should be expanded as much as possible. Making information goods available for free also has positive distributive implications as it avoids the large monopoly rents that otherwise accrue to the holders of intellectual property rights.

However, the second fundamental problem is that when it comes to commercializing products, private companies are frequently superior to publicly funded entities. For example, Steve Jobs was probably better at

designing iPhones than DARPA would have been. Even if DARPA wanted to, it could not contract out the design of goods that have not yet been imagined to visionaries such as Steve Jobs. This limits the spheres in which the efficient outcome can be achieved via public investment in information goods.

In all areas where we rely on private actors for commercialization, we are thus left with second-best policy options that involve granting some monopoly power to the private creators of information goods by awarding them intellectual property rights. When dealing with second-best policy options, everything is about trade-offs: Although granting limited monopoly power may be desirable to provide incentives to innovate, the level of monopoly power that we currently provide and the resulting monopoly rents seem far in excess of the cost of investment in a number of industries, as indicated by record profits. Since the resulting rents extract surplus from consumers to the benefit of large corporations, with undesirable distributive implications, we should counteract them. One avenue is to weaken intellectual property rights and the associated monopoly power; another avenue is to tax away some of the rents earned by corporations, e.g. by charging them licensing fees for the publicly created technologies that they rely on.

Increasing returns that stem from network externalities are an additional factor that is very relevant in the context of digitization and information goods. The greater the number of existing users on a digital platform such as Facebook, Google or Amazon, the more attractive the platform becomes for new users. This makes platforms even stronger natural monopolies, amplifying the associated rents and superstar effects as well as the resulting inequality. Since these network effects frequently revolve around data, we can reduce the power of such natural monopolies by giving consumers more freedom in how their data is used and by forcing interoperability between different platforms via standards for data exchange. For example, if consumers can grant a start-up that they trust access to the same social network graph, search history or shopping history that established internet firms already have, the monopoly power of existing corporations would be curbed.

The Rise of Artificial Intelligence

Digitization and superstar firms are just the beginning of a larger wave of technological progress that will be of increasing relevance going forward and that centers on the rise of Artificial Intelligence (see e.g. Agrawal et al., 2019; Korinek and Stiglitz, 2019; Acemoglu and Restrepo, 2019). Traditionally, when we were concerned about inequality, we have been thinking of inequality between different groups of humans, such as workers and capitalists, and how they compete for scarce resources. This is based on the anthropocentric notion that only humans consume final goods – a notion that has been perfectly reasonable for much of the history of mankind.

At the present stage, we humans still feel mostly in control of the intelligent algorithms and machines that we have created and that we interact with on a daily basis. However, to an objective observer, things look a little bit different than they used to a few decades ago: Artificially intelligent agents (AIAs) play an increasingly important role in our economy and are in fact more and more in control of us humans. A growing number of corporate decisions that affect us are made by AIAs – from screening job applicants to providing loans. A growing number of our personal decisions is strongly influenced (or, one might say, manipulated) by AIAs – ranging from what we read and buy to whom we date and how we vote. AIAs also act increasingly autonomously in our economy, for example engaging in financial transactions or driving on our roads.

From a broader perspective, humans and intelligent machines are both entities that share certain basic economic properties: first and foremost, they both absorb scarce resources. These resources serve to meet their maintenance needs and ensure their survival. Although the absorption basket of the two types differ – for example, humans consume bread whereas machines absorb electricity – the basic economic function is the same. Furthermore, both types of entities also supply their factor services to the economy – human labor or machine labor, and they both follow defined laws-of-motion.

In a recent paper on “The Rise of Artificially Intelligent Agents” (Korinek, 2018), I observe that – as we are entering a period in which ever more intelligent machines surpass the capabilities of humans in a growing number of areas – competition over scarce resources

may increasingly play out between humans and artificial entities. The most tangible present manifestation of such entities are high-tech corporations. They absorb a growing share of the economy’s resources – for example, the human labor they employ, the raw materials that go into computing and data centers, and the electricity they consume (server farms absorb close to 10% of the world’s electricity production, by some estimates). They also accumulate rising levels of wealth. And although they are notionally owned by humans, the de-facto level of control exerted by their owners is rather low. As AIAs gain ever more autonomy, their actions increasingly surprise their human creators and owners and are frequently misaligned with the objectives of their human owners, as for example Mark Zuckerberg experienced when he found out about the role of Facebook in recent elections. From this perspective, the question of “who owns intelligent machines or algorithms” is increasingly irrelevant – ownership without control is meaningless. The true masters of the universe, as Silicon Valley refers to the founders of the largest and most influential high-tech corporations, are not so much the humans who own them but the algorithms themselves.

One of the prime challenges for humanity in the age of AI will be to ensure that humans will continue to prosper and obtain a fair share of the resources produced by our shared human-AIA economy. The themes and policy proposals of the preceding sections of this policy note take on even greater urgency when viewed through this lens: Undoing technological redistribution and steering technological progress are even more important when they are about the distribution of resources between humans and artificial entities. Moreover, reducing the monopoly power of digital superstars gains extra importance when it is about maintaining the consumption share of humans in our common economy.

In spite of all these measures, human labor may well become irrelevant in the labor market in coming decades (Korinek and Stiglitz, 2018, 2019). Satisfying the basic needs of us humans would then require income from sources other than labor, whether they be labeled a social dividend, an allocation of subsistence income, or a universal basic income. The political difficulty of direct handouts can be reduced by providing many of the services that we humans rely on, such as healthcare and education, for free. Furthermore, some may view it desirable to subsidize humans to perform tasks that provide meaning, even though they are wasteful and redundant from an economic perspective. The stark

alternative would be to let Malthusian forces play out, which would lead to large unnecessary suffering in a world of growing abundance.

Anton Korinek is an Associate Professor of Economics at the University of Virginia. Contact: anton@korinek.com

Endnotes

* I would like to thank Susan Helper, Rebecca Henderson, Ioana Marinescu and Dani Rodrik for their thoughtful comments and Hasan Toprak for excellent research assistance. Special thanks go to João Oliveira for sharing data to calculate the human capital share of income.

References

- Acemoglu, Daron and Pascual Restrepo (2019), “The Wrong Kind of AI? Artificial Intelligence and the Future of Labor Demand,” Working Paper.
- Agrawal et al. (2019), *The Economics of Artificial Intelligence*, NBER and University of Chicago Press.
- Autor, David (2015), Why are there still so many jobs? The History and Future of Workplace Automation, *Journal of Economic Perspectives* 29(3).
- Brinca, Pedro, João B. Duarte and João G. Oliveira (2019), “Investment-specific Technological Change, Taxation and Inequality in the U.S.” *MPRA Paper 91463*, University Library of Munich, Germany.
- Korinek, Anton (2018), “The Rise of Artificially Intelligent Agents,” Working Paper.
- Korinek, Anton and Ding Xuan Ng (2018), “Digitization and the Macroeconomics of Superstars,” Working Paper.
- Korinek, Anton and Joseph E. Stiglitz (2019), “Artificial Intelligence and Its Implications for Income Distribution and Unemployment,” forthcoming in Agrawal et al.: *The Economics of Artificial Intelligence*, NBER and University of Chicago Press.
- Korinek, Anton and Joseph E. Stiglitz (2018), “Worker-Replacing Technological Progress,” Working Paper.

Election Law and Political Economy

Ethan Kaplan

Turnout

Early Voting: Voter turnout as a percentage of the voting-age population in the United States was 56% in the 2016 election (<http://www.pewresearch.org/fact-tank/2018/05/21/u-s-voter-turnout-trails-most-developed-countries/>). Though turnout was high compared to other recent U.S. elections, it was the sixth lowest among the 32 OECD (Organization for Economic Cooperation and Development) countries for which voting age population is available. Only Luxembourg, Slovenia, Poland, Chile, Latvia, and Switzerland had lower turnout rates. Moreover, turnout in non-Presidential federal elections is often around 40%. The countries with the highest voter turnout are Belgium (87%), Sweden (83%) and Denmark (80%).

In contrast to most developed countries, federal elections in the United States do not fall on weekends nor are they national holidays. This makes it difficult for voters to make it to the polls in order to cast a ballot. Voting on a working day is a likely contributor both to the overall low turnout rates in the United States as well as to the substantially higher turnout of higher income Americans. Those earning more than \$150,000 per year vote at a 50% higher rate in presidential elections and at a 100% higher rate in midterm elections than those making less than \$5,000 per year. More generally, turnout is roughly monotonically increasing in income. This is unsurprising since low income workers have less flexible work schedules (Enchautegui, 2013).

One policy which many states have used to increase turnout is early voting. Currently, 32 states have mandated or allowed polling stations to open before election day for voters to cast ballots in person.

Minnesota has the most generous early voting system with 46 days of early voting including openings on weekends and weekday nights. It also had the highest turnout rate of any state in the country in 2016.

Kaplan and Yuan (forthcoming) estimate the impact of expanding access to voting using an early voting homogenization law in Ohio from 2012. They find that expanding early voting increases turnout by 0.22 percentage points per day of early voting. This implies that a state like Minnesota yields an additional 10 percentage points of turnout from its 46 days of early voting. Moreover, Kaplan and Yuan (forthcoming) show that the effect is 20% higher for women than for men. The effects are also higher for those of child-bearing and working age. Additionally, they find higher impacts on political Independents and Democrats. Other research on the impact of early voting in Florida has shown higher impacts on African-American voters than on whites (Herron and Smith, 2012). Simulations of the creation of a national early voting law at the level of Minnesota suggest that such a policy would have increased turnout, made the electorate more representative of the population, and altered the outcome of the 2016 Presidential election and the majorities in the Senate in both 2012 and 2016.

Alternate (or additional) possible legal changes that would likely have a similar impact would be to declare election day a national holiday or to place it on a weekend. This also would likely increase turnout and make the electorate more representative of the voting population.

Photo Identification: A number of laws have been passed in recent years which restrict access to the ballot box.

Two examples of these are (1.) photo identification laws which require voters to present certain forms of official forms of identification and (2.) felon disenfranchisement laws which disallow felons and former felons from voting. Ethnic minorities and young voters are most likely to show up to polling stations without a voter id. Both of these groups are under-represented in the national electorate. Citrin, Green and Levy (2012) show that informing voters about new voter id laws has a 1 percentage point higher impact than just encouraging voters to turn out. Thus, photo id laws do both reduce turnout and make the electorate less representative of the population. Laws preventing incarcerated felons or former felons no longer in jail from voting also have similar impacts.

The reasons given for photo identification laws are that they protect elections from voter fraud such as double voting. However, photo identification laws, as discussed above, have substantial negative impacts upon voter turnout and tilt the electorate towards over-represented voters. There is recent evidence on voter fraud and find it to be minimal at best. Goel, Meredith, Morse, Rothschild and Shirani-Mehr (2017) find that, in contrast, to the sizable negative impact of photo id laws on voting, double voting accounts for at most 0.02% of all votes and almost all of these are consistent with small amounts of measurement error in the voter files. Getting rid of photo id laws would improve the representativeness and fairness of the electoral system as would elimination of felon disenfranchisement laws.

Voter Registration: Though only a slim majority of voters participate even in Presidential elections, voters are required to register to vote in order to vote. Only North Dakota does not require voter registration. Moreover, most states require registration well in advance of the election. In the 2016 Presidential elections, 86.8% of registered voters cast ballots. Thus, a substantial portion of the barrier to full participation in elections is a barrier from registration requirements. Moreover, this problem is exacerbated by the practice by election commissions who manage the voter rolls of purging the rolls of inactive voters. Voters with absences in participation across multiple federal elections are often dropped from voter registration databases. In most states, when on election day, these voters then go to vote, they are unable to do so because election day registration is not allowed.

Ansola-behere and Konisky (2006, Political Analysis) show that the introduction of statewide voter registration (some counties already had pre-existing registration systems in place) in New York and Ohio in 1965 and 1977 respectively depressed turnout by 10 percentage points. In some states, voters may register to vote and vote on election day. This is called election day registration. Burden (2011) shows that election day registration lowers the registration turnout penalty by 2 percentage points. Braconnier, Dormagen and Pons (2017) show similar reductions in voter turnout from registration laws in France. A recent paper by Cantoni and Pons (2018) shows that election day registration has the single largest impact upon turnout of any election policy which varies across states in the United States. In Ohio, voters must be registered 28 days or more in advance of the election in order to vote on election day. There is some overlap between the period of early voting and the mandatory registration period so that voters can register to vote and vote early in the same location at the same time. Kaplan and Yuan (forthcoming) show that the effect upon turnout of same day registration days is much larger than for normal early voting days.

In recent years, some states have created automatic voter registration systems where voters are automatically registered by the state government at age 18. Oregon became the first automatic voter registration state in 2016. Rhode Island and Vermont have now followed Oregon and the state of Washington has passed legislation to enact automatic voter registration as of 2019. Since these laws are new, not much analysis has been done on their impact. However, they represent one possible way to increase voter turnout. Overall, many laws in the United States discourage full participation and differentially make voting difficult for lower income families and minorities who tend to have less flexible jobs.

Gerrymandering

In the United States, representatives to the House of Representatives are elected in winner-take-all districts. Every ten years, following the census, these districts are redrawn for the purposes of equating representation across districts. This is done at the state level. However, particularly when there is unified government, the party in power shapes the districts in order to maximize their party's seat shares in the state and federal legislatures.

For example, in Wisconsin in 2012, the Republicans gained 60 out of 99 seats (almost 2/3) with 48.6% of the vote share. Nationally, in 2014, the Democrats in the House of Representatives won 48.8% of the national vote against 47.6% for the Republicans; however, the Republicans won 234 seats and the Democrats won 201. The Brennan Center for Justice estimates that Republicans had 16 to 17 additional seats in the House of Representatives due to partisan gerrymandering during the 115th Congress (2017-2018). In other words, in the absence of partisan gerrymandering, the Democratic party would have had majority control in the House of Representatives in the 115th Congress.

Though gerrymandering is not necessarily biased towards over-representation of higher income and white voters, in recent years, it has been more heavily used by the Republican party. This is due to two factors: (1.) more progressive states often pass laws which make gerrymandering more difficult and (2.) the Republican party won a record number of state legislatures and governorships in 2010 which then allowed them to gerrymander in a large number of states. A recent paper by Jeong and Shenoy (2017) estimates that obtaining majority control over the legislature and thus the gerrymandering process increase the probability of majority victory in a given district in the first election after redistricting of 11 percentage points.

In 14 states, including the large state of California, the legislature has shifted redistricting decisions from elected governors and legislatures to independent commissions. In 6 cases, these are non-partisan commissions and in 8 cases, commissions are bi-partisan. In addition, measures have been developed to assess the imbalance in partisan representation across districts. The “efficiency gap” measure computes the number of wasted votes (votes above majority) as a fraction of total votes cast in a district. With modern data, we have both policies to implement non-partisan or bi-partisan gerrymandering as well as the ability to assess its success. Possible remedies for gerrymandering include national legislation or state-level legislation establishing non-partisan or bi-partisan redistricting commissions. These are currently 21 state with bi-partisan or non-partisan commissions. Alternatively, states could pass legislation electing representatives in proportion to state-level vote shares for Congressional seats.

Campaign Contributions and Political Influence

Election laws in the United States allow for monetary contributions from private individuals and from corporations. Courts have interpreted campaign contributions as protected under freedom of speech provisions of the first amendment to the U.S. Constitution. Contributions have historically been capped due to the strong state interest in reducing the possibility of “influence” or the “perception of influence”, either of which the courts have claimed could be destabilizing for democracy. However, what do we know about the actual impacts of campaign contributions?

First of all, the amount of monetary expenditures is large and rapidly increasing. In the year 2004, political campaigns for presidential candidates spent \$850 million, campaigns for House elections spent \$660 million, and Senate campaigns spent \$550 million. By 2016, those numbers had risen to \$1.5 billion of spending by Presidential campaigns, \$970 million by House campaigns, and \$670 million by Senate campaigns. However, in addition to campaign expenditures, independent expenditures by individuals and corporations have become substantially more important in the past decade. Two important decisions by federal courts (*Citizens United v. FEC*, 2010 and *Free Speech v. FEC*, 2013) allowed for unlimited independent expenditures by individuals or corporations in favor of political candidates. In the wake of these decisions, independent expenditures rose from \$330 million in 2008 to \$1.0 billion in 2012 and \$1.4 billion in 2016.

But do campaign expenditures have an impact? Most political campaigns spend the vast majority of their funds on political advertisement and “Get Out The Vote” (GOTV) operations¹. Gerber and Green (2015) have done field experiments on the impacts of GOTV operations. In particular, they have randomized contact with voters in various formats: canvassing (door knocking), phone calls, mailings and even text messages. They then use official voter registration data to check turnout for those whom they contact and those whom they do not. They find positive impacts upon turnout for all forms of contact. More recently Vincent Pons analyzed the effects of a massive randomized voter persuasion effort (5 million doors were knocked) by the Socialist Party in France in the 2012 French

Presidential elections. The socialist party randomized its canvassing efforts. It differs from the Gerber and Green experiments in that the object was not just voter turnout but also voter persuasion. Pons (2018) did not find a statistically significant impact on voter turnout; however, he did find a 3 percentage point increase in the vote share for the Socialist Party in the precincts which were randomly canvassed. In addition, he found that some of the persuasive effects of the initial canvassing experiment persisted to future elections. So campaign expenditures on GOTV operations is highly effective.

There has also been recent very convincing research done on the impact of political advertisement. Spenkuch and Toniatti (2018) estimate the impact of political advertisements by looking at turnout and vote share differentials across media market boundaries within states where there was a differential partisan composition of advertisements. Similar to Pons (2018), Spenkuch and Toniatti (2018) fail to find a statistically significant impact upon voter turnout though the estimated effect is positive. However, they do find that an additional ten ads in favor of a candidate raises that candidate's vote share by 0.22 percentage points. Surprisingly, these results are estimated for Presidential campaigns where one would expect communication overload to have driven the impact of marginal impact of advertisement to zero.

Most campaign expenditures, therefore, have sizable impacts upon election outcomes. Thus contributors to campaigns, either through direct contributions, through Political Action Committees, or through independent expenditures, influence politics and policy outcomes. There are two main ways in which campaign expenditures can influence political outcomes. They can influence who is elected ("election motive") and they can influence how politicians vote once elected ("influence motive")². Either way contributors exert undue influence over the political system.

Who are the contributors? Particularly with the rise in independent expenditures, contributors are increasingly the wealthiest members of society. Since the Citizens United ruling, over 20% of contributions have been from the top 0.01% of income. Moreover, the fraction of contributions coming from the top 0.01% of the income earners has been rising over time even since the passage of Citizens United (Bonica et al., 2013). These contributions likely both support more conservative candidates (even within the Democratic party) and

also make candidates more conservative particularly on economic issues. A new paper by Kaplan, Spenkuch and Yuan (2018) show that members of Congress are much more likely to support an interest group's position on a bill when they have received a contribution from the interest group. Moreover, they find that when votes take place on high news pressure days (i.e. when there is likely to be little coverage of the bill in the press), the responsiveness of members of Congress to campaign expenditures increases by up to a factor of 10. This shows that interest groups not only influence who gets elected but also how politicians vote once elected. The skewness in the earnings distribution of contributors thus likely pushes policy, and particularly economic policy, strongly towards low levels of taxation and government expenditures. It is not surprising that as campaign contributions have increased at very rapid rates over the past 5 decades, tax rates on income and particularly on capital income have declined so dramatically.

Finally, the fact that individuals, corporations and interest groups can spend so much money on political campaigns means that they can threaten to spend money if candidates do not vote in accordance with their policy desires. With partisan gerrymandering, it is increasingly difficult for wealthy individuals or interest groups to threaten candidates in general elections. However, in highly partisan districts, there is often room for a primary challenge to an elected politician. Chamon and Kaplan (2011) write a theoretical model where interest groups or individuals can contribute to a campaign or threaten to contribute to an opposing campaign. Direct evidence on threats is difficult to find since such threats are technically illegal. Chamon and Kaplan (2011) provide empirical evidence in support of their model of threats. However, there is also anecdotal evidence. For example, Grover Norquist's Americans for Tax Reform has been successful in getting almost all Republicans to sign a pledge that they will not raise taxes. These tax pledges received press coverage during the Obama-era government shutdowns when legislators were threatened with primary challenges for agreeing to expenditure increases. Also, the Senate Conservatives Fund, headed by Ken Cuccinelli, threatened to recruit and fund primary challenges to any Republicans not voting for repeal of the Affordable Care Act. The ability for individuals and corporations to spend unlimited amounts of money in support of a candidate in the post-Citizens United era increases the influence of money on

politics and policy while simultaneously increasing the difficulty in detecting the impact of money.

Many countries limit campaign expenditures. In many countries, government provides funding for campaigns in lieu of private funding. Some countries such as Canada not only put limits on individual contributions but also put limits on aggregate expenditures by individual campaigns. Well-crafted empirical work on campaign contribution limitations are rare. This is because campaign contribution limitations are usually decided at the country level. This makes causal estimation of the impact of such laws difficult. Recent work by Avis et al. (2017) show that campaign contribution limitations in Brazil led to greater political competition and entry by candidates with lower average income and wealth. The United States could tighten campaign contribution limits to individual campaigns, to political parties and could pass a constitutional amendment limiting or banning independent expenditures.

Ideological Influence

Political advertising influences voters but so do the media and so do think tanks and membership organizations. In the 19th century, newspapers were explicitly partisan. This died out towards the end of the 19th century (Gentzkow and Shapiro). Starting in the late 1980s, conservative talk radio reintroduced partisan news on a large scale. This was followed by the introduction of the conservative Fox News Channel (FNC) in 1996. FNC soon became the most popular news channel on cable television. It was clearly much further to the right than other news organizations (Martin and Yurukoglu). Studies have used the staggered introduction of Fox News as well as randomness in the channel order of Fox News across towns to estimate the impact of Fox News on voting patterns. DellaVigna and Kaplan (2007) and Martin and Yurukoglu (2016). They both find a positive impact on the Republican vote share from the expansion of FNC of around 0.5 percentage points in the 2000 Presidential elections. They find similarly sized effects in other races (Senate, House of Representatives). Martin and Yurukoglu (2016) also cover later years and find that as FNC expanded, its influence grew to 6.3 percentage points in the 2008 elections. These are extremely large effects. Martin and McCrain (2018) also find a substantive conservative shift and increase in viewership in local news due to consolidation by the

Sinclair Media Group.

What policies exist to potentially mitigate the impact of conservative news? Fox news slants their news strongly in favor of the Republican party (as MSNBC does in favor of the Democratic party). Other countries such as the United Kingdom have laws mandating minimum levels of impartiality. Moreover, the United Kingdom has purely publicly financed television (the BBC). Impartiality laws regulating particularly TV media as well as funding for a public television news station could help restore the imbalance created by FNC. As an example, in the U.K., a regulatory agency (Ofcom) reviews all forms of media for content and mandates corrections for misleading or untrue information. Their power comes from the ability to levy fines. In extreme cases, Ofcom (Office of Communications) can deny operating licenses or deny mergers and acquisitions.

In addition to the media, membership organizations potentially wield political influence and political power. As income and wealth inequality have risen, organizations have proliferated supporting the interests of business owners: the Chamber of Commerce has expanded and become a partisan organization and wealthy donors have formed other organizations (i.e. the American Enterprise Institute, the Cato Institute, and the Heritage Foundation). Think tanks also exist which support the Democrats or are on the political left (i.e. the Center for American Progress, the Economic Policy Institute). However, the organizations supporting business owner interests are unsurprisingly substantially better funded. The main set of organizations which have historically represented the interests of working people have been labor unions. However, unions have been in strong decline over the past 45 years. With a peak of 35% representation in the private sector, unionization rates in the private sector are now below 7% (Naidu, 2019). During this time period of declining unionization, we have also seen large reductions in top marginal tax rates from 94% in the late 1940s to 70% in the 1960s to the current 35%. Interestingly, the conservative move in economic policy has not been mirrored in social policy suggesting a potential role for unions.

Moreover, the conservative supreme court has through multiple decisions (Harris v. Quinn, 2014; Janus v. AFSCME, 2018) eroded the ability of public sector unions to collect fees from their members. This will further reduce the size and power of public sector unions going forward.

Though we do not have a credible estimate of the impact of unions directly on policies, recent work by Hertel-Fernandez, Feigenbaum and Williamson (2018) estimate the impact of unionization upon the Democratic vote share. They compare changes in county-level presidential vote shares in states when a right to work law is passed to neighboring counties in other states that never passed a right-to-work law. Right to work laws ban contracts which allow unions to charge dues to all workers. Right-to-work laws have been shown to have very large impacts on state private sector unionization rates. Hertel-Fernandez et al. (2018) find no difference in right-to-work versus non-right-to-work law state presidential vote shares before right-to-work laws are passed. However, immediately following passage, they find a decline in the Democratic vote share of 3.5% which expands over the following decade to 15 percentage points. Thus strengthening labor law by banning Right-to-work laws, allowing for public sector union due collection, and streamlining the union election process (i.e. allowing for a card check process) would also help lower income individuals express themselves politically at state and national levels. Naidu (2019) discusses in greater policies and ideas for increasing the strength of the union movement in the United States.

Summary

In sum, political institutions in the United States favor higher income individuals over lower income individuals and ethnic majorities over ethnic minorities. This is accomplished through a myriad of policies which impact who votes, allow for differential influence and access by the wealthy, structure voting districts to dilute the impacts of under-represented voters, and allow for oversized influence of pro-business owner ideas through media and membership organizations.

There are many policies, some implemented in other countries, which could help restore greater balance to political competition and ideological competition. Some policies which we have discussed in this piece which would help equalize the political process include:

Turnout

- (1.) Expand early voting, particularly weekend voting and same day registration.
- (2.) Make election day a national holiday.
- (3.) Get rid of photo identification laws.
- (4.) Eliminate voter registration or at least introduce automatic voter registration.
- (5.) Eliminate felon disenfranchisement laws.

Monetary Influence

- (1.) Limit campaign contributions to very small amounts such as \$50 per individual per candidate per election or eliminate private funding and go to a public funding system.
- (2.) Ban or strongly limit independent expenditures.
- (3.) Ban expenditures for corporations.

Gerrymandering

- (1.) Mandate Independent Redistricting Commissions

Ideological Influence

- (1.) Disallow politically biased television news.
- (2.) Make union representation easier.

Ethan Kaplan is an Associate Professor of Economics at the University of Maryland. Contact: kaplan@econ.umd.edu

Endnotes

- 1 Information on campaign and independent expenditures come from [opensecrets.org](https://www.opensecrets.org), which synthesizes data from the Federal Election Commission.
- 2 Grossman and Helpman (2002).

References

- Ansolabehere, Stephen and David Konisky (2006), “The Introduction of Voter Registration and Its Effect on Turnout”, *Political Analysis* 14(1), pp. 83-100.
- Avis, Eric, Claudio Ferraz, Frederico Finan, and Carlos Varjao (2017), “Money and Politics: The Effect of Campaign Spending Limits on Political Competition and Incumbency Advantage”, *NBER Working Paper*.
- Bonica, Adam, Nolan McCarthy, Keith Poole and Howard Rosenthal (2013), “Why Hasn’t Democracy Slowed Rising Inequality?”, *Journal of Economic Perspectives* 27(3), pp. 103-124.
- Braconnier, Celine, Jean-Yves Dormagen and Vincent Pons (2017), “Voter Registration Costs and Disenfranchisement: Experimental Evidence from France”, *American Political Science Review* 111(3), pp. 584-604.
- Burden, Barry and Jacob Neihsel (2011), “Election Administration and the Pure Effect of Voter Registration on Turnout”, *Political Research Quarterly* 66(1), pp. 77-90.
- Cantoni, Enrico and Vincent Pons (2018), “Does Context Trump Individual Drivers of Voting Behavior? Evidence from U.S. Movers”, working paper.
- Chamon, Marcos and Ethan Kaplan (2013), “The Iceberg Theory of Campaign Contributions: Political Threats and Interest Group Behavior”, *American Economic Journal: Economic Policy* 5(1), pp. 1-31.
- Citrin, Green and Levy (2012)
- DellaVigna, Stefano and Ethan Kaplan (2007), “The Fox News Effect: Media Bias and Voting”, *Quarterly Journal of Economics* 122(3), pp. 1187-1234.
- Enchautegui, Maria (2013), “Nonstandard Work Schedules and the Well-Being of Low Income Families”, *working paper*.
- Gentzkow, Matthew, and Jesse M. Shapiro (2010), “What Drives Media Slant? Evidence from US Daily Newspapers”, *Econometrica* 78(1), pp. 35-71.
- Gerber, Alan and Donald Green (2015), **Get Out the Vote: How to Increase Voter Turnout**, Brookings Institution Press, Washington, D.C.
- Goel, Sharad, Marc Meredith, Michael Morse, David Rothschild, and Houshmand Shirani-Mehr (2017), “One Person, One Vote: Estimating the Prevalence of Double Voting in U.S. Presidential Elections”, working paper.
- Grossman, Gene and Elhanan Helpman (2002), *Special Interest Politics*, MIT Press.
- Herron, Michael and Daniel Smith (2012), “Souls to the Polls: Early Voting in Florida in the Shadow of House Bill 1355”, *Election Law Journal* 11, pp. 331-347.
- Feigenbaum, James, Alexander Hertel-Fernandez, and Vanessa Williamson (2018), “From the Bargaining Table to the Ballot Box: Political Effects of Right to Work Laws”, working paper.
- Jeong, Dahyeon and Ajay Shenoy (2017), “Political Capture: the Case of U.S. Congressional Redistricting”, working paper.
- Kaplan, Ethan, Jorg Spenkuch and Haishan Yuan (forthcoming), “News, Information and Representation”, *American Economic Journal: Applied Economics*.
- Kaplan, Ethan and Haishan Yuan (2018), “Early Voting Laws, Voter Turnout, and Partisan Vote Composition: Evidence from Ohio”, revision request at *American Economic Journal: Applied Economics*.
- Martin, Gregory and Josh McCrain (2018), “Local News and National Politics”, revision requested at *American Political Science Review*.
- Martin, Gregory and Ali Yurukoglu (2017), “Bias in Cable News: Persuasion and Polarization”, *American Economic Review* 107(9), pp. 2565-2599.
- Naidu, Suresh (2019), “Worker Collective Action in the Mid-21st Century”, working paper.
- Pons, Vincent (2018), “Will a Five-Minute Discussion Change Your mind? A Countrywide Experiment on Voter Choice in France”, *American Economic Review* 108(6), pp. 1322-1363.
- Spenkuch, Jorg and David Toniatti (2018), “Political Advertising and Election Results”, *Quarterly Journal of Economics* 133(4), pp. 1981-2036.

Using wage boards to raise pay

Arindrajit Dube

Growing wage inequality and nature of wage setting

During the past 40 years, the United States has experienced a sharp and sustained rise in wage and income inequality. The high level of inequality in the United States reflects both a disconnect between (1) average wages and productivity, and (2) top and bottom wages.

As shown in Figure 1, much of the growth in labor productivity has gone to growth in wages at the top of the distribution. Panel (A) of figure 1 shows the growing gap between median compensation and average productivity into the gap between capital and labor share, compensation inequality, and differential price growth for consumer and producer baskets. While net

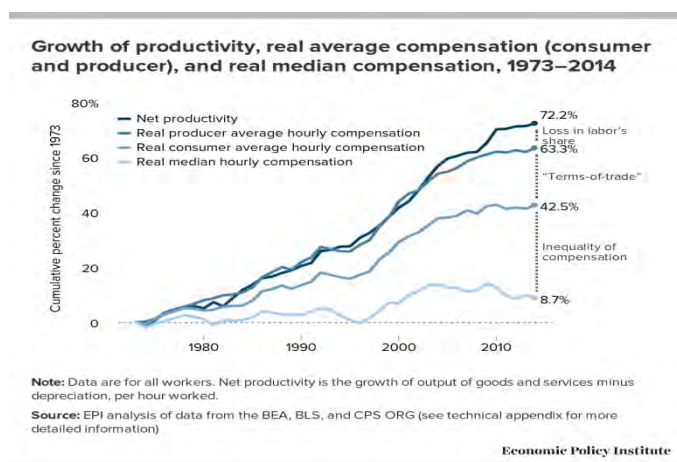
productivity grew by 72% between 1973 and 2014, over that period median real compensation grew only by 8%.

Much of the gap between mean productivity and median compensation arises from growing inequality in the labor market that has grown steadily over this period, especially since 1980. This is reflected in the fact that mean compensation grew by around 43% over this period, much more than median.

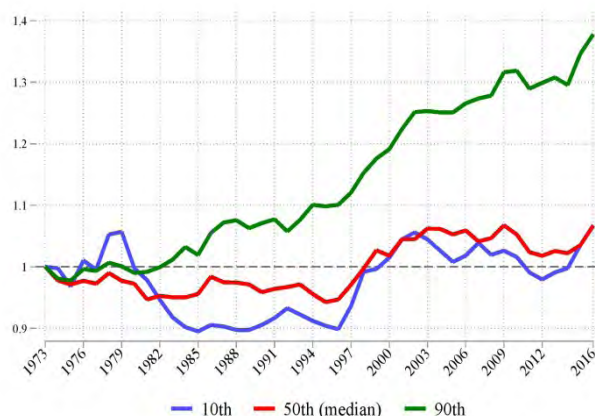
The pattern of inequality is also reflected in panel (B) which shows that the 90th percentile real wage grew by over 35 percent between 1973 and 2016, while the median and 10th percentile real wage grew by approximately 6 percent over the same period. While there are some divergences in the bottom of half of the wage distribution, they are small compared to the sizable and growing gap in pay between those at the top compared to the rest of the workforce.

Figure 1 Trends in wages and productivity

Compensation and productivity (Index=1 for 1973)



Wages by Percentiles (Index=1 for 1973)



Finally, besides the inequality in wages, labor's share itself has fallen since 2000, suggesting a smaller pie to be shared among wage earners. This likely puts additional downward pressure on wages.

While globalization and technological change have likely played a role, a sizable body of evidence in economics suggests institutions have been important contributors to these trends as well—including collective bargaining and statutory minimum wages. The stagnation of the federal minimum wage since the 1980 contributed to real wage declines at the bottom (Autor et al. 2016), and the erosion of collective bargaining led to wage declines in the middle (Farber et al. 2018).

Moreover, economic theory and evidence increasingly point to the importance of labor market power, suggesting that the *laissez faire* equilibrium without collective bargaining may be better understood as being monopsonistic rather than perfectly competitive to a first order approximation. Growing evidence suggests the importance of firm's wage policies in explaining pay differences across workers (e.g., Song et al. 2018; Card et al. 2015). Recent evidence using matched employer-employee data suggests that around 20% of the variance in log wages is explained by firm specific factors. This is consistent with the older institutionalist tradition that highlighted “wage contours” (e.g., Dunlop 1957) – though it would probably be called “frictional wage inequality” today. It is also consistent with a large body of evidence that employers have substantial power to set wages without the ironclad discipline of labor market competition (Manning 2003). Recent work has provided high quality evidence on employers' power to set wages in a range of sectors from retail to online platforms (e.g., Dube, Naidu, Jacobs and Suri 2018; Dube, Giuliano and Leonard 2018; Cauldwell and Oehlsen 2018). Moreover, there is evidence that the extent of concentration in a local labor market is correlated with the level of wages (Azar et al. 2017; Benmelech et al. 2018).

Taken together, the weight of evidence suggests that we have moved from a labor market in the U.S. that was based on labor market negotiations via collective bargaining to one where employers increasingly have power to set wages subject to limited labor market discipline. Moreover, in the mid 20th century, collective bargaining likely served as a reference point for wage setting in the non-union sector. Erosion of these norms likely weakened pressures on non-union employers who, today, are freer to exercise their monopsony

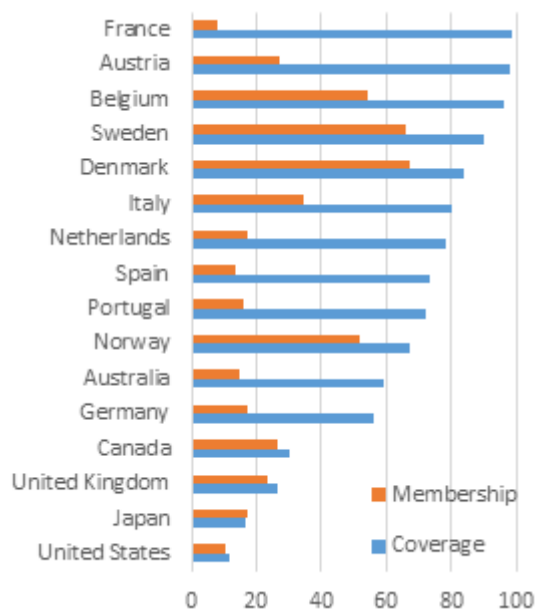
power. In other words, the weakening of labor market institutions and attendant norms that historically provided countervailing power can explain the extent and nature of real wage stagnation for most working Americans.

Role of labor market institutions

In the era following the second world war, the key countervailing force in the U.S. labor market came from unions. Overall union membership reached a height of around 35 percent of the workforce in the mid 1950s. Unions affected wages both directly as well as indirectly through pattern bargaining as in the “Treaty of Detroit” (Levy and Temin, 2011). However, since then, union membership has steadily fallen, and stands at around 12 percent today (under 7 percent in the private sector).

The impact of a falling union membership has been particularly acute due to the enterprise-level bargaining structure in the U.S. (and other countries like UK and Canada), which differs greatly from countries like France, Germany, Australia where collective bargaining *coverage* (share of jobs covered by collectively bargained contracts) is much greater than the union *membership* rates.

Figure 2 Union membership and coverage rates



Source: OECD Stats.

France, for example, has an 8 percent union membership rate (similar to the U.S.). And yet, over 95% of its workforce is covered by extensions of nationally negotiated collective bargaining contracts. While coverage rates also have fallen across the developed world over the past several decades, the outcomes have varied greatly between (1) countries where membership and coverage rates have remained more stable (e.g., Ghent system countries like Denmark), (2) countries where extension of contracts/sectoral bargaining has kept coverage rates high even as membership rates have fallen (e.g., France), and (3) countries with enterprise-level bargaining where membership *and* coverage rates have both fallen sharply (e.g., U.S.). Overall, this decline in union density has likely led to substantial reductions in wages in the middle of the distribution (Farber et al. 2018; see also Suresh Naidu's chapter in this volume).

In the U.S., efforts towards reversing the union decline have focused on making it easier to organize (e.g., Employee Free Choice Act). Suresh Naidu's chapter in this volume discusses innovative strategies to ease the ability of workers to self-organize. Such changes would be valuable indeed. However, when starting from a 10% density (7% in private sector), even with greater organizing ease, it would take decades to rebuild union density back to anywhere near the post-war high water mark. In addition, enterprise level bargaining also incentivizes employer opposition; while sectoral bargaining means competitors are subject to the same wage bargain, enterprise level bargaining may put an individual employer at a greater competitive disadvantage. Finally, most of the efforts to reform labor law to make it easier to organize have failed in the political arena.

In contrast, more success has come on the front of setting minimum wage standards, with state-level changes taking the lead. These policy successes have occurred while mainstream economics has engaged in a rethink about the costs and benefits of minimum wage policies fueled by credible empirical evidence, coupled with increasing theoretical understanding of search frictions and other sources of labor market power. In a comprehensive analysis, Cengiz et al. (2018) show that in the US, there is little evidence of either overall job loss or losses for lower skilled groups for minimum wages that are up to half the median wage for full time workers. This level of minimum wage is consistent with other OECD countries and US experience in the 1960s and 1970s.

Recent “bolder” minimum wage experiments (e.g., California, Massachusetts, New York) have set forward a path to \$15/minimum wage by 2023. Some of the “bolder” policies will lead to a substantially greater share of workers (>25%) whose wages will be directly or indirectly tied to the minimum wage, and will exceed the range of minimum-to-median wage ratios we have seen to date. Therefore, empirical evidence will need to assess whether these policies go “too far.” However, even if a national \$15 minimum wage is found to be a beneficial policy, the minimum wage would have to be much higher to substantially move the median wage, requiring massive compression and possible risks (reduced hiring and business investment, increased price inflation). Obviously, there are serious limits to using a single policy lever to affect the entire wage distribution.

The Wage Board Approach

One alternative to a single, high minimum wage involves instituting a wage board that sets multiple minimum pay standards by sector and occupation, potentially chosen using consultation with stakeholders, such as business and worker representatives and elected representatives (Andrias 2016, Madland 2018). The use of sector and occupation allows particular job types to have minimum pay standards. This would allow raising wages not just for those at the very bottom, but also for those at the middle. For example, janitors working for building services contractors are typically paid fairly low wages. However, the pay ranges are typically above statutory minimum wages, and over 90% of workers in this sector make above the federal minimum wage. This means the minimum wage is not a particularly effective tool to raise pay in this sector. The wage board approach can better reach such low (but not the lowest) rungs of the pay scale. This is effectively done in countries where there are extensions of collective bargaining contracts, but can also be done by setting multiple minimum pay levels statutorily.

An example of a wage board approach comes from Australia, which has a combination of 1) national minimum wage, 2) the “Modern Awards” system of industry and occupation-specific minimums, 3) enterprise-level collective bargaining. Around 36% of the workforce is covered by collective bargaining contracts, but another 23% are covered by awards only.

The awards are set by a federal tribunal whose members are appointed by the government to serve until the age of 65. Most of these are by industry, although some (e.g., nurses or pilots) are by occupation. There are 122 such awards, and within each there are a host of wage rates based on skill-requirements or experience; there may be anywhere between a handful to several dozen pay grades specified in each agreement.

As a practical matter, the annual wage increases are largely similar each year for most awards and pay grades. For example, in 2018, most Modern Award pay standards were increased by 3.5% in low-wage sectors like Retail or Hospitality, the same amount as the baseline minimum wage increase. In some years or particular cases, there may be some additional adjustment to wages to further boost pay in the lowest categories of work to achieve greater pay compression.

While it is difficult to definitively assess the impact of the Australian system of labor standards, broad metrics offer a positive verdict. Household inequality in Australia is more muted as compared to the U.S. with a 90/10 ratio of 4.3 instead of 6.3 according to most recent data from the OECD. Importantly, as shown in Figure 3 below, the median wage has kept up with the mean wage in Australia much more than in the US, where the median has stagnated since the 1980s. This evidence is consistent with the view that labor market institutions like the Modern Awards system helped ensure broad

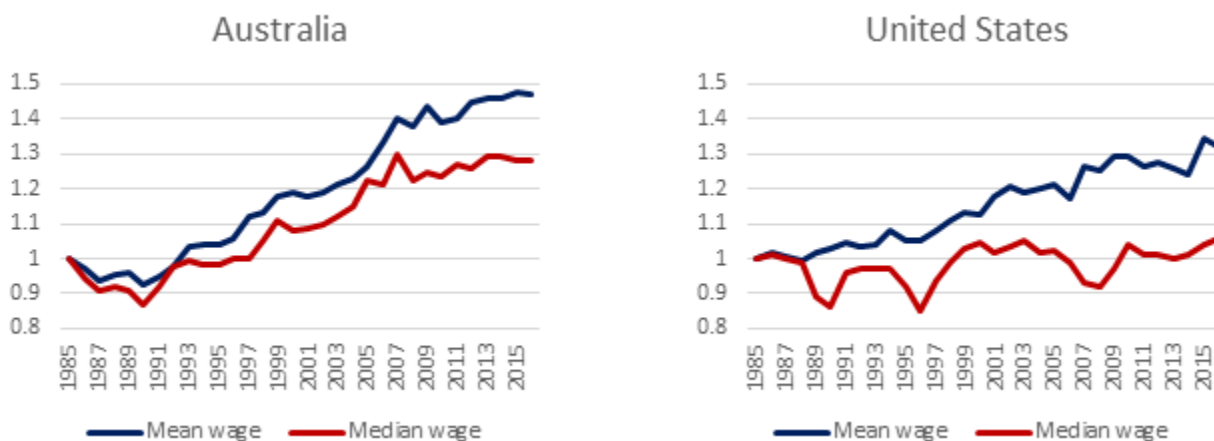
based prosperity and ameliorating the growth in wage inequality in Australia, when compared to the US benchmark of de-unionization and erosion of wage standards.

At the same time, the more muted growth in inequality is not associated with any obvious differences in labor market performance. While the current unemployment rate in Australia is 5.3% as opposed to 3.9% in the US, averaged over the past 10 years, the respective figures are 5.5% in Australia and 6.9% in the US. Moreover, focusing on younger or lower skilled workers does not yield very different comparisons. Overall, the Australian evidence is broadly consistent with the perspective that judiciously applied wage setting using a wage board system can help ameliorate wage inequality without causing any serious harm to the labor market.

Application to the United States

In order to institute wage boards at the national level in the U.S., federal law would need to be changed. However, at the state level, at least 5 states (Arizona, Colorado, California, New Jersey, and New York) already have legislation on the books that allows for constituting wage boards by industry or occupations. At the same time, these boards have been used infrequently. Most prominently, they were used to raise the overall minimum wages in California in the 1990s, and more

Figure 3 Evolution of mean versus median wages – Australia and US



Source: OECD Stats. Notes: wages are PPP-adjusted real wages, indexed to 1985 value.

recently to establish a fast food minimum wage in New York. However, there been little effort to use the wage board mechanism to target wages for the middle of the distribution.

At the same time, the machinery is in place to push for a broader array of wage standards. State experimentation with wage boards to set standards higher up in the wage distribution—as in the Australian case—could play a possibly useful role in mitigating wage stagnation and inequality. Moreover, other states can follow suit and establish similar wage board legislation to those in place in California.

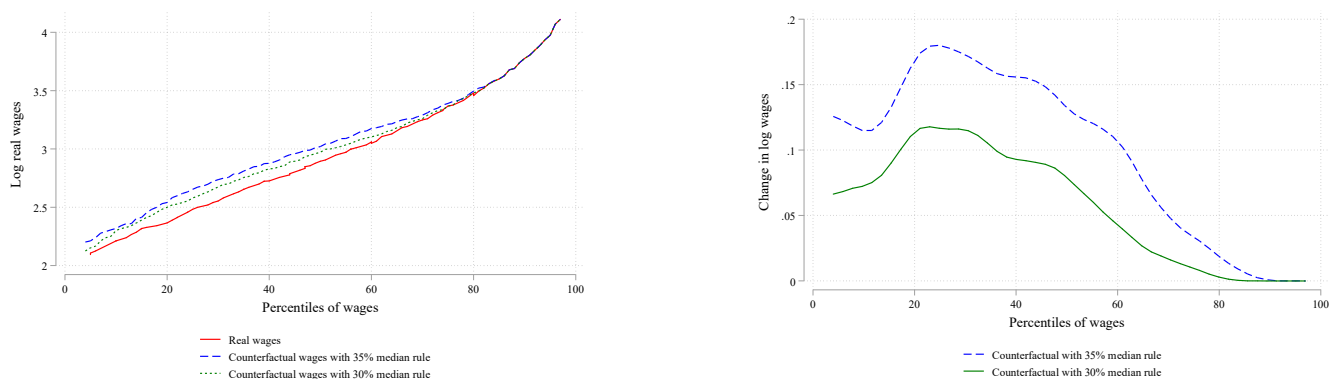
While details can vary, a wage board system would set minimum pay standards by sector and occupation. This allows the mechanism to affect the distribution of wages not just at the very bottom but additionally toward the middle of the distribution. As an illustration, I simulate the effect of a wage board by imposing region-by-industry-by-occupation standards, separately calculated by region (specifically 9 census divisions), 17 two-digit industries, and 6 occupational groups producing a total of 102 wage standards. The choice of standards is of course a key issue: to show how this may affect wage inequality, I consider two standards: in the first (“low”) I set the minimum to 30% of the median wage in each of the 102 categories in that particular Census division, while in the second (“high”) I set it to 35% of the median. While as a share of the median wage these two standards seem not to be very far apart, they do imply quite different bites for the policy, as I show below.

As a starting point, the wage standards would be binding for 20% and 31% of workers under the low and high scenarios, respectively. In other words, the low and the high scenarios straddle the Australian case—where around 23% of workers’ wages are set by the modern award system. However, Australia also has a substantially higher set of workers with collectively bargained wages (36%) than in the US (12%). Therefore, the “high scenario” would still imply a smaller same set of workers who are covered by either collective bargaining or by a wage board as in Australia.

As shown in Figure 4, overall, both the high and low scenarios imply substantial wage gains, especially for the bottom and middle of the wage distribution. Under the low scenario, the 20th, 40th and 60th percentile wage rises by 13, 9 and 4 percent, respectively. When we consider the higher scenario the wage gains extend somewhat further: wages at the same percentiles would rise by 19, 15 and 12 percent, respectively. It is useful to contrast these distributional impact of wage boards with those from typical minimum wage increases in the U.S, which mostly fade out by the 20th percentile of the wage distribution. In other words, wage boards are much better positioned to deliver gains to middle-wage jobs than a single minimum pay standard.

Of course, these calculations are illustrative and make many simplifying assumptions such as ruling out additional spillover effects, changes in composition of jobs, to name a few. However, what they show is that a suitably chosen wage standard can substantially raise middle and bottom wages and lower wage inequality.

Figure 3 Effect of hypothetical wage boards on log wages



Risks and Challenges

If a wage board system leads to substantial increases in wages, it is reasonable to be concerned about possible unintended consequences. It is important to acknowledge all experimentation involves unknowns, especially when it involves substantial changes to the wage structure. Two natural sources of concern would be price increases and impact on jobs. This suggests it is useful to structure any changes incrementally, giving us enough time to learn from the experiments.

What could such steps look like? First, it might be useful to pilot wage boards for some specific sectors, ideally ones which would be well targeted to moving wages towards the bottom and middle of the distribution. Second, when taking the wage board to scale (i.e., applying to a broad set of sectors), it makes sense to make pay increases gradual, preventing any sharp increases. Especially when it comes to the national level, broad based wage increases can affect aggregate demand, as well as price inflation which is closely monitored by the Federal Reserve. A full-fledged wage board system works best when the monetary and wage board authorities work in partnership to maintain stable wage and price growth with periodic adjustment to wage compression. The experience in Australia, as well as many countries with national-level sectoral bargaining, suggests this is feasible.

Of course, having the mechanism in place is not a guarantee that it will be effective. At the same time, we need more arrows in our quiver to tackle income inequality and wage stagnation. And wage boards may well be one of those.

Arindrajit Dube is a Professor of Economics at the University of Massachusetts Amherst. Contact: adube@econs.umass.edu

References

- Andrias, Kate. 2016. “The New Labor Law,” *The Yale Law Journal* 126(1): 1–261.
- Autor, David, Alan Manning, and Christopher Smith. 2016. “**The contribution of the minimum wage to US wage inequality over three decades: a reassessment.**” *American Economic Journal: Applied* 8(1), 58–99.
- Azar, José, Ioana Marinescu, and Marshall I. Steinbaum. 2017. “Labor market concentration.” *National Bureau of Economic Research Working Paper* No. w24147.
- Benmelech, Efraim, Nittai Bergman, and Hyunseob Kim. 2018 “Strong employers and weak employees: How does employer concentration affect wages?” *National Bureau of Economic Research Working Paper* No. w24307.
- Card, David, Jörg Heining, and Patrick Kline. 2013. “Workplace heterogeneity and the rise of West German wage inequality.” *The Quarterly journal of economics* 128(3): 967–1015.
- Cauldwell, Sydnee and Emily Oehlsen. 2018. “Monopsony and the Gender Wage Gap: Experimental Evidence from the Gig Economy.” Working paper.
- Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2018. “The Effect of Minimum Wages on Low-Wage Jobs: Evidence from the United States Using a Bunching Estimator.” *CEP Discussion Paper* No 1531.
- Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri. 2018. “Monopsony in Online Labor Markets.” *American Economic Review Insights* (forthcoming).
- Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard. 2018. “Fairness and frictions: The impact of unequal raises on quit behavior.” *American Economic Review* (forthcoming).
- Farber, Henry S., Daniel Herbst, Ilyana Kuziemko, and Suresh Naidu. 2018. “Unions and Inequality Over the Twentieth Century: New Evidence from Survey Data.” *National Bureau of Economic Research Working Paper* No. w24587.
- Levy, Frank and Peter Temin. 2011. “Inequality and Institutions in 20th Century America.” Pp. 357–86 in *Economic Evolution and Revolution in Historical Time*, edited by J. Rosenbloom, P. Rhode, and D. Weiman. Palo Alto, CA: Stanford University Press.
- Madland, David. 2018. “Wage Boards for American Workers.” Center for American Progress Report, available at <https://www.americanprogress.org/issues/economy/reports/2018/04/09/448515/wage-boards-american-workers/>
- Manning, Alan. 2003. *Monopsony in Motion*. Princeton University Press.
- Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2018. “Firming up inequality.” *Quarterly Journal of Economics* (forthcoming).

An Expanded View of Government's Role in Providing Social Insurance and Investing in Children

Sandra E. Black and Jesse Rothstein

A key issue in economic policy is determining which goods or services the government should provide — either by producing them directly or by funding others to do so. Traditional economic theory suggests that, if markets are functioning properly, competitive market forces will generate efficient provision of goods without intervention from the government. But the conditions required for this efficiency result are quite strong, and there are a great many goods and services to which they do not apply. When they do not, there is no *a priori* reason to think that welfare will be higher when the good is left to the private market than when it is publicly provided.¹

We argue for a larger public role in the provision of two categories of goods that long experience has shown are drastically under-provided by private markets, for which theory and evidence each clearly indicate that public provision would improve welfare. Government should do more to support families in the raising and educating of children. It should also play a larger role in insuring against certain types of risks that individuals and families face.

We first discuss childhood investments, including expenses relating to child care and early education as well as post-secondary education. A key characteristic of these investments is that the costs are lumpy and are, to a large extent, incurred by parents, though most of the benefits accrue to the children. These services are

also very expensive, are tightly concentrated in time, and in many cases occur decades before the benefits are realized. In addition, as we discuss below, research has shown that there are many public externalities associated with these investments. Efficient private provision, even assuming agency problems and externalities could be addressed, requires borrowing large amounts against the child's future earnings. Such loans are not available on the private market, and even when the government can create a market for them (as for student loans), issues of asymmetric information create moral hazard problems in the absence of stringent government regulation. Together, these three issues ensure that without a larger role for government in providing and/or funding these services we will see dramatic underinvestment in child development.

The second category we discuss is insurance. There are many types of insurance that are efficiently provided by the (often carefully regulated) private market, such as car or homeowners' insurance. However, other types of insurance that individuals would highly value either are not available at all on the private market or are available only at extremely high prices to a small share of the group that could benefit. These include insurance against job loss in recessions, against illness, against outliving one's savings, and against long-term care expenses. We focus on these because these risks detract importantly from individual welfare, and, in the absence of insurance, lead to quite costly responses such as over-saving for

retirement and going without needed health care. It is not likely that private insurance markets can be made to function well in these areas, even with aggressive regulation. Instead, these are natural candidates for public provision.

We want to emphasize that we are not the first to propose public provision of the types of services considered here. Indeed, many childhood investments are already publicly provided. Most obviously, we provide free public K-12 education, and most college students attend public institutions, albeit often with substantial and growing tuition bills. Similarly, we also provide public insurance against unemployment, longevity (via Social Security), and unforeseen medical costs. Our purpose is merely to articulate a common intellectual justification for these programs, and to advocate for expanding them to cover important needs that are not currently covered.

Child care and education

It is increasingly recognized that high-quality child care and early childhood education, prior to entering kindergarten, is a key investment with important implications for children's long-run outcomes (e.g., Elango et al. 2016; Deming 2009). This investment has important impacts on others – a more educated child may benefit the rest of society through reduced reliance on public support, productivity spillovers, and reduced criminal activity – and one can make a strong case for public intervention on the basis of these externalities alone.

But even if we set aside potential impacts beyond the immediate family, there is little reason to expect private decisions to be optimal. The decision to purchase child care and early childhood education, along with the burden of paying for them, rests on parents, while it is the children's futures that are at stake. This type of principal-agent problem could generate inefficient investment – parents may invest less in children's education than the children themselves would, could their future selves be allowed to control the decision and bear the costs (Brown et al., 2012). Society has addressed this issue for children between the ages of 6 and 18 through the funding and provision of public education. However, public investments are heavily tilted toward older children: President Obama's Council of Economic Advisers estimated that, in 2015, combined annual local,

state, and federal expenditure per child was 63% higher for those between 6 and 11 than for those between 3 and 5 (Council of Economic Advisers 2016b).

Importantly, this early childhood period also coincides with a period when parents' own earnings are lower and they are thus least able to afford the substantial expenses that young children bring. Because the benefits of investments in child development are not easily collateralized (as they appear through better longer-run outcomes far in the future, when the children are grown, and as children's own future earnings cannot be encumbered), it may be difficult for parents to borrow to finance this investment, even if they wanted to (Caucutt and Lochner 2012).

These factors point to a need for government intervention. A range of existing programs help parents when children are young—the Head Start pre-school program is one example (Deming 2009) —but these programs are relatively small and tightly targeted to the very poor. The high expenses of early childhood are a burden not just for the poor but also for middle class families, who are at similar risk of under-investment. Broader based support, beginning with funding for high quality childcare and pre-K, would help ease this burden and ensure that children receive the appropriate investments when they are small. This funding should be accompanied by careful and thorough quality regulation, as research has shown the important benefits of high quality childcare and pre-school; lower quality programs are less effective. (Council of Economic Advisers 2016a, Chapter 4)

Higher education presents its own set of challenges. Children are much more involved in their own postsecondary education decisions, which at least partially aligns the benefits with the decision-making. However, financing this investment remains difficult—again, as the benefits are realized far in the future in terms of higher earnings for the students, private markets have trouble funding the loan without extensive government involvement. This, combined with evidence of positive externalities generated by education (Moretti 2004), suggests a need for government intervention.

We currently support a credit market for higher education by guaranteeing student loans and enforcing repayment (for example, by making the loans non-dischargeable in bankruptcy). But this creates its own problems. Institutions vary widely in quality in

ways that are hard for students to perceive, and many institutions take loans on students' behalf without providing education of commensurate quality (Deming et al. 2012). The inability of the government to effectively regulate quality has led to what some have termed a student loan crisis—student debt ballooned during and after the Great Recession, in large part due to increasing attendance at low-quality institutions that are unlikely to yield a positive return.

With debt financing, students bear the risk of regulatory failures, as they must repay their loans even if the institution they attended turns out not to have helped them succeed. There is little indication that this risk leads to better decisions, given the lack of information available to students. Welfare could be increased if the government simply insured against poor outcomes by reducing the costs of a college education. This could take any of a number of forms, ranging from increased public provision through growth of the public higher education sector (with restrained or eliminated tuition made up through additional investment of tax revenue) to larger public grant programs built on the existing Pell Grant. Given the evidence that much of the student debt problem is concentrated among individuals who never completed their degrees (CEA 2016c), one attractive solution is to make the first two years of college free.² In addition to removing enormous leverage and risk from individual portfolios, this would also resolve an agency problem, creating appropriate incentives for government to regulate quality aggressively.

Social Insurance

Risk is a fundamental part of life and can have major welfare consequences. Without insurance, even prosperous families would lack economic security, as they could risk being thrown into poverty at any time. Insurance can protect them from this risk and thereby promote security.

Many private insurance markets – for example, auto insurance, life insurance, and homeowners insurance –work fairly well, in part due to extensive regulation. But people face many risks for which there is no private insurance market. Economists have long studied the information and other failures that prevent well-functioning insurance markets. There are many types of market failures, but three are common:

- **Moral hazard.** People with insurance are cushioned against the negative consequences of bad outcomes, and may not act prudently to minimize their risk if someone else will bear the cost. Someone with car insurance may be more willing to park a car in a dangerous neighborhood, raising the risk that it could be stolen. This raises the likelihood of damage, making it very expensive to insure truly unavoidable risks.
- **Heterogeneity in risk and adverse selection.** Risk often varies across people in predictable ways – some may have a genetic propensity toward a particular disease, for example. When the potential purchaser of insurance has more information about his or her type than the insurer, the insurer must price insurance under the assumption that the consumer is of the high-risk type, setting the premium very high. This makes it difficult or impossible for those with low risk to buy insurance at any reasonable price, leaving many people uninsured. Even when risk type is verifiable, so that the insurer is able to sell insurance to both types, those with above-average risks can face very high prices, which may price them out of the market and reduce overall welfare. When insurance markets do exist, insurers have strong incentives to design policies narrowly, excluding preexisting conditions or denying coverage after the fact for costs that the individual had hoped to have covered, leaving the insured with less protection against risk than he or she would like.
- **Common/correlated risks.** The insurance business relies on risks being random but total cost being predictable: While we can't know whose house will burn down in the coming year, we can estimate how many fires there will be, and the fire insurer can thus plan on predictable expenses. Some risks, however, are not idiosyncratic, but shared: Although a major earthquake may occur only once every 50 years, when it does a large share of policyholders will file claims. If the insurer does not maintain a large reserve fund, it will not have enough funds to pay claims when they arrive. Much insurance regulation centers around managing this problem, often unsuccessfully (Sjostrom 2009; Coffee Jr. 2011). The problem is much worse for risks associated with the business cycle, such as unemployment, where payout events tend to occur

when capital is hardest to obtain, and this type of insurance is very expensive to provide.

These pathologies can cause insurance markets to break down, leaving many people uninsured. But it is important to remember that their presence, and the resulting absence of a private insurance market, is *not* a signal that the welfare value of insurance is low; people would prefer to insure themselves against consequential risk, but they find themselves unable to do so at a fair price.

Where adverse selection or correlated risks cause insurance markets to fail, there is a role for the government. By mandating the purchase of insurance at a price that is fair on average, or by providing the insurance out of general revenues, the government can generate risk coverage despite information asymmetries that would otherwise cause insurance markets to fail due to adverse selection.³ Similarly, the federal treasury has unique access to credit markets during recessions, when it typically faces very low – often zero – borrowing rates. This enables it to insure business cycle risks that private insurers cannot cover. (The government generally has no special ability to manage moral hazard, so where that is the primary source of failure there may not be a public role.)

These arguments are not new – there has long been a recognized role for the public sector in the provision of social insurance (Hacker, 2008). An example is disability insurance: those who become disabled often lose their livelihoods, but information asymmetries make it very difficult to purchase disability insurance on private markets. Where insurance is available, underwriters often require extensive medical exams, and refuse to write insurance for those with preexisting conditions. Through the Social Security Disability Insurance and Supplemental Security Insurance programs, the government provides insurance to all, providing monthly income to the disabled through (in the case of SSDI) premiums levied on the healthy. This necessarily entails some transfer from the healthy to the sick, which may be socially desirable in its own right. But it also enables everyone to obtain protection from a very serious risk to their livelihood, protection that would not otherwise be available.

But the scope of existing social insurance programs is tightly circumscribed. There are many important risks that reduce welfare that are not covered, even though

the insurance market failures are similar to those that are. An expanded scope would be welfare improving (Hacker 2008). We review several leading candidates for new or expanded social insurance coverage below.

First, however, we address a common counterargument. It is often suggested that moral hazard makes it unwise to expand social insurance: With coverage, people will take on risks that they would not otherwise. There are two responses to this. First, in several of the cases that we discuss below, the degree of risk is well outside the individual's control, so there is no scope for moral hazard. We are not concerned that longevity insurance will lead people to live longer than is efficient, for example, or that business cycle insurance for workers will lead them to court recessions. Second, and most importantly: There is nothing specific to *social* insurance that creates or magnifies moral hazard, which is equally likely to afflict holders of private insurance policies. This may limit the extent of insurance that it is prudent to provide, but absent some reason to expect moral hazard to be a particular problem in a case under consideration, it does not constitute an argument against the provision of insurance.

We next discuss several specific risks that are presently difficult to insure, where a new or expanded government role would increase welfare.

Unemployment insurance

A worker who loses her job will need to finance consumption, often for an extended time and generally with very limited ability to borrow against future income, until she finds a new job. In the absence of insurance, workers must maintain large savings against this possibility. There is little efficiency benefit of forcing people to bear the portion of this risk that reflects changes in their employers' prospects outside of their control. Accordingly, a joint federal-state program has since 1935 provided unemployment insurance to workers, financed by payroll taxes. To limit coverage to the true risks and avoid moral hazard, payments are limited to those laid off from their jobs; those who quit or are fired for cause are generally not eligible. Another measure taken to limit moral hazard is time limits on benefits, typically 26 weeks, which are meant to encourage aggressive job search before the benefits run out. The idea here is to balance moral hazard against the need for insurance (Bialy 1978; Chetty 2008; Schmieder

and von Wachter 2016).

This insurance is quite valuable to workers. But the uniform program described above does not address systematic differences in job availability. The chance that someone will be able to find a job by the end of 26 weeks of benefits, even with diligent search, varies enormously over the business cycle, with much higher rates of benefit exhaustion in recessions. The federal government often extends benefits in recessions, but these extensions are ad hoc and often come too late to be helpful. Expanding the existing system to automatically extend benefits when the economy weakens could significantly improve worker welfare (Landais et al., 2018a,b).

Such a policy offers two other benefits, beyond the basic risk protection. First, unemployment benefits act as Keynesian demand stabilizers, boosting consumption among those with high propensities to spend. Automatic extensions would ensure that that boost arrives when the economy needs the additional demand, not afterward. Second, the moral hazard argument for encouraging active job search by limiting the duration of benefits is much attenuated in recessions, when there are more job searchers than jobs and there would be little efficiency cost to reduced search effort among the unemployed.⁴

Old age insurance

Someone who reaches old age without adequate savings has no good options. At that point, it is too late to go back to work, so the only choices are to sharply reduce consumption or to rely on transfers from family members.

If lifespans and investment returns were predictable, the private solution would be straightforward: People would set aside money during their working lives to consume during retirement. However, the unpredictability of lifespans and of investment returns creates substantial risk, and the only way people can protect themselves is by saving much more than they will likely need. This precautionary savings reduces welfare (Lusardi 1988; Abel 1995; Hubbard 1987; Kotlikoff et al. 1986; Mitchell et al. 1999).

There is little efficiency benefit of making people bear either of these risks. People would be much better off with insurance that guaranteed them a stable income as long as they lived, and there is no real moral hazard

problem here. Annuities are financial products that provide insurance against these risks. However, annuity markets generally function poorly, for reasons that are not entirely understood but include adverse selection and the substantial complexity of the products on offer (Brown et al. 2008; Finkelstein and Poterba 2004; Brown et al. 2008). Annuities are generally priced well above their actuarial value, and few people buy them (Lockwood 2012; Mitchell et al. 1999).

The resulting market failure creates a clear public need that has long been recognized. Since 1935, Social Security Retirement Insurance has provided a mandatory retirement annuity to all American workers, with some modest redistribution from high- to low-earners and from younger to older cohorts. But Social Security was designed to be just one part of what was intended as a three-legged stool, with private pensions and individual savings providing the other legs. In the last decades, one of these legs has nearly disappeared – only one-fifth of full-time private-sector workers are covered by a defined-benefit pension.⁵ And the last leg never functioned well – less than 60% of those approaching retirement have any retirement savings, and over half of those have less than \$100,000 saved – not enough to last through a long retirement. Those who do save still face uncovered risks from financial market volatility as well as the risk of outliving their savings.

Given these failures, there is a strong case for expanding the Social Security annuity, growing it to cover a much larger share of expected retirement consumption. Again, this would not distort private decision-making, except for those few who would prefer to dramatically reduce their consumption in retirement. To recognize this heterogeneity, one might combine an expansion of required Social Security with an optional public annuity, structured as an option to top up one's Social Security benefits through voluntary additional contributions. This would be a much less risky way to save for retirement than via 401(k)s, and with even modest take-up would be welfare improving.

Health insurance

Medical care absorbs an ever-growing share of national expenditures, due in large part to incredible progress in medical science that makes it possible to treat a wide range of ailments that were previously untreatable, though often at a high cost. This makes health insurance

essential to economic security. However, health insurance markets rarely work well, for two reasons: First, information asymmetries and adverse selection are particularly salient in health insurance, and insurance companies often deny insurance to those at high risk (for example, those with preexisting conditions). Second, the correct care is not easily observable and often depends on doctor discretion, creating a principal-agent problem that leads to extensive controls on care usage, frequent denial of coverage for legitimate claims, and enormous hassles for the policyholder. The end result is that, absent very strict regulation, many go without insurance altogether, and those who have insurance still face large risks of being bankrupted by charges that the insurance company denies (Swartz 2003; Himmelstein et al. 2000, 2009; Warren et al., 2000). In addition to the missing markets for those with the highest risk, there is another important case for government insurance. Health insurance and health care generates positive externalities to society, in that healthier citizens are more productive and those who do not have health insurance may ultimately depend on safety net programs when they get sick (Finkelstein et al. 2018; Brown et al. 2015).

We have already created a successful public health insurance program for the elderly, the hardest group to cover due to the high and predictable costs of aging. For those who do not yet qualify for Medicare, we rely primarily on private insurance markets, which must be extensively regulated. Regulators often fail to keep the markets functioning well, and even with careful regulation many remain uncovered. Public provision of health insurance would dramatically reduce bureaucracy and transaction costs in health care while ensuring universal coverage.

Long-term care insurance

Related to health insurance is long-term care insurance. Many people need labor-intensive personal care at the end of their lives, and following major illnesses. This is a major expense that arrives when people are least able to accommodate it. In principle, healthy people could save against possible future long term care needs, but the range of possible outcomes is so enormous, and the variability so high, that few do this. Instead, people fall back on public welfare programs – stories abound of people intentionally exhausting their assets to qualify

for Medicaid – or on their families, who are often called upon to provide enormous amounts of uncompensated care for which they are not well qualified.

Insurance against this major risk of old age would be extremely valuable and would not create serious moral hazard problems (Cohen et al. 2018). Private insurance markets exist, but are complex, hard to understand, and badly undersubscribed. Adverse selection seems to combine with individual optimization failures – people prefer to wait until they are sick to purchase long term care insurance, but at that point it is unaffordable (Brown and Finkelstein 2007, 2009; Brown et al. 2008). The existence of Medicaid as a fallback option also makes it more feasible to go without this type of insurance, though the low quality of Medicaid care leaves a large uncovered risk (Brown and Finkelstein 2008, 2011). Finally, this type of insurance creates serious time consistency problems, similar to but more severe than those in traditional health insurance: Insurers have strong incentives to collect premiums from the healthy, then deny coverage and/or change the terms of coverage as people approach the need to make a claim.

The net effect is that people face large uncovered risks, while the existence of the public program for the poor, via Medicaid, means that the government still winds up bearing much of the cost. Providing higher quality, universal public coverage, without requiring recipients to exhaust all assets before using it, would be expensive, but would much improve welfare (Cohen et al. 2018).⁶

Conclusion

While private provision of goods often yields the efficient outcome, there are a number of goods that are not efficiently provided in the private market. Here, we have outlined two such situations—investments in child care and education, and insurance against risks created by business cycles, poor health, and old age. Because private markets work poorly for these goods, and the costs of market failure are large, standard economic reasoning implies a significant role for government provision. The reduction in economic insecurity that this would bring could help to improve political stability as well, by reducing the stakes that people perceive in discussions of trade, immigration, technological change, and countercyclical policy (Inglehart and Norris, 2016). Many observers (e.g, Hacker, 2018) have

pointed to economic anxiety as a potential contributor to populist reactions in the U.S. and many European countries; a public sector that acts to reduce the risk that households face could ameliorate this, generating political spillovers and improving the state of the country more broadly.

Sandra Black is a Professor of Economics and holds the Audre and Bernard Rapoport Centennial Chair in Economics and Public Policy at the University of Texas at Austin. Contact: sblack@austin.utexas.edu

Jesse Rothstein is a Professor of Public Policy and Economics at the University of California, Berkeley. Contact: rothstein@berkeley.edu

Endnotes

- 1 This discussion focuses on efficiency rather than distribution. Concerns about inequality may provide their own motivation for government intervention, even in the absence of market failures.
- 2 A version of this was recently proposed by Congressman Bobby Scott of Virginia with his “Aim Higher” proposal that would include two years of free community college.
- 3 In addition to creating an opportunity to obtain insurance where none otherwise exists, government-provided insurance can also avoid many of the pathologies of private insurance in the presence of adverse selection, such as aggressive underwriting, preexisting condition exclusions, and ex post denial of apparently valid claims.
- 4 Another possible form of insurance against job loss would be through wage insurance. Evidence suggests that individuals who lose their jobs often suffer wage losses in their next jobs (Davis and von Wachter, 2012). Wage insurance would offset some of these losses. Wage insurance was included in President Obama’s 2017 Budget Proposal.
- 5 <http://www.epi.org/files/2013/epi-retirement-inequality-chartbook.pdf>
- 6 There was an attempt to do this via the Affordable Care Act, but the program was under-resourced and was eventually abandoned. The problem has not gone away, and welfare would be improved if we did this right.

References

- Abel, A. B. (1985). Precautionary saving and accidental bequests. *The American Economic Review*, 75(4), 777-791.
- Baily, M. N. (1978). Some aspects of optimal unemployment insurance. *Journal of public Economics*, 10(3), 379-402.
- Brown, D., Al Kowalski, and I. Lurie. (2015). Medicaid as an investment in children: What is the long-term impact on tax receipts? National Bureau of Economic Research working paper 20835.
- Brown, J. R., Casey, M. D., & Mitchell, O. S. (2008). Who values the Social Security annuity? New evidence on the annuity puzzle. National Bureau of Economic Research working paper 13800.
- Brown, J. R., & Finkelstein, A. (2007). Why is the market for long-term care insurance so small?. *Journal of Public Economics*, 91(10), 1967-1991.
- Brown, Jeffrey R., and Amy Finkelstein. 2008. "The Interaction of Public and Private Insurance: Medicaid and the Long-Term Care Insurance Market." *American Economic Review*, 98 (3): 1083-1102.
- Brown, J. R., & Finkelstein, A. (2009). The private market for long-term care insurance in the United States: a review of the evidence. *Journal of Risk and Insurance*, 76(1), 5-29.
- Brown, Jeffrey R., and Amy Finkelstein. 2011. "Insuring Long-Term Care in the United States." *Journal of Economic Perspectives*, 25 (4): 119-42.
- Brown, J. R., Kling, J. R., Mullainathan, S., & Wrobel, M. V. (2008). Why don't people insure late-life consumption? A framing explanation of the under-annuitization puzzle. *American Economic Review*, 98(2), 304-09.
- Brown, M., Karl Scholz, J., & Seshadri, A. (2011). A new test of borrowing constraints for education. *The Review of Economic Studies*, 79(2), 511-538.
- Caucutt, E. M., & Lochner, L. (2012). Early and late human capital investments, borrowing constraints, and the family. National Bureau of Economic Research working paper 18493.
- Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of Political Economy*, 116(2), 173-234.
- Coffee Jr, J. C. (2011). Systemic risk after Dodd-Frank: contingent capital and the need for regulatory strategies beyond oversight. *Columbia Law Review*, 111, 795.
- Cohen, Marc, Judith Feder, and Melissa Favreault (2018). "A New Public-Private Partnership: Catastrophic Public and Front-End Private LTC Insurance." Unpublished manuscript, UMass Boston LeadingAge LTSS Center and Urban Institute, January.
- Council of Economic Advisers (2016a). *Economic Report of the President*.
- Council of Economic Advisers (2016b). "The disconnect between resources and needs when investing in children." Issue brief.
- Council of Economic Advisers (2016c). "Investing in Higher Education: Benefits, Challenges, and the State of Student Deb." Issue brief.
- Davis and von Wachter (2012). "**Recessions and the Costs of Job Loss**" Brookings Papers on Economic Activity.
- Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3), 111-34.

- Deming, David J., Claudia Goldin, and Lawrence F. Katz. 2012. “The for-profit postsecondary school sector: Nimble critters or agile predators?” *Journal of Economic Perspectives*, 26 (1): 139-64.
- Elango, S., A. Hojman, J.-L. García, and J.J. Heckman (2016). “Early childhood education.” In Moffitt, R. (Ed.), *Economics of Means-Tested Transfer Programs in the United States II*. Chapter 4, pp. 235-298. Chicago, IL: University of Chicago Press.
- Finkelstein, A., N. Mahoney, and M. Notowidigdo. (2018). What does (formal) health insurance do, and for whom? *Annual Review of Economics*, 10: 261-86.
- Finkelstein, A., & Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the UK annuity market. *Journal of Political Economy*, 112(1), 183-208.
- Hacker, Jacob (2008). *The Great Risk Shift: The New Economic Insecurity and the Decline of the American Dream*. Oxford University Press.
- Hacker, Jacob (2018). The great risk shift that helped deliver Donald Trump. *The American Prospect*, October 24.
- Himmelstein, D. U., Warren, E., Thorne, D., & Woolhandler, S. (2005). Illness and injury as contributors to bankruptcy. *Health Affairs*, 24(1), 570.
- Himmelstein, D. U., Thorne, D., Warren, E., & Woolhandler, S. (2009). Medical bankruptcy in the United States, 2007: Results of a national study. *The American Journal of Medicine*, 122(8), 741-746.
- Inglehart, Ronald, and Pippa Norris. Trump, Brexit, and the rise of populism: Economic have-nots and cultural backlash. HKS Faculty Research Working Paper Series RWP16-026, August 2016.
- Kotlikoff, L. J., Shoven, J., & Spivak, A. (1986). The effect of annuity insurance on savings and inequality. *Journal of Labor Economics*, 4(3, Part 2), S183-S207.
- Landais, C., Michaillat, P., & Saez, E. (2018). A macroeconomic approach to optimal unemployment insurance: Theory. *American Economic Journal: Economic Policy*, 10(2), 152-81.
- Landais, C., Michaillat, P., & Saez, E. (2018). A macroeconomic approach to optimal unemployment insurance: Applications. *American Economic Journal: Economic Policy*, 10(2), 182-216.
- Lockwood, L. M. (2012). Bequest motives and the annuity puzzle. *Review of Economic Dynamics*, 15(2), 226.
- Lusardi, A. (1998). On the importance of the precautionary saving motive. *The American Economic Review*, 88(2), 449-453.
- Mitchell, O. S., Poterba, J. M., Warshawsky, M. J., & Brown, J. R. (1999). New evidence on the money’s worth of individual annuities. *American Economic Review*, 89(5), 1299-1318.
- Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1-2), 175-212.
- Schmieder, J. F., & Von Wachter, T. (2016). The effects of unemployment insurance benefits: New evidence and interpretation. *Annual Review of Economics*, 8, 547-581.
- Sjostrom Jr, W. K. (2009). The AIG bailout. *Washington & Lee Law Review*, 66, 943.
- Swartz, K. (2003). Reinsuring risk to increase access to health insurance. *American Economic Review*, 93(2), 283-287.
- Warren, E., Sullivan, T., & Jacoby, M. (2000). Medical problems and bankruptcy filings. *Norton’s Bankruptcy Adviser*.

Towards a Better Financial System

Anat R. Admati*

Introduction

A healthy and stable financial system enables efficient resource allocation and risk sharing. A reckless and distorted system, however, causes enormous harm. The cycles of boom, bust, and crisis that repeatedly plague banking and finance are symptoms of deep governance and policy failures. Reinhart and Rogoff (2009), who studied financial crises over many years and jurisdictions, conclude that crises are preventable but that governments are themselves part of the problem, either because they mishandle their own finances and borrow too much, or they fail to prevent recklessness by households and firms.

Despite efforts at regulatory reforms since the 2007-2009 financial crisis, too little has changed. The rules governing the financial system remain complex, inadequate and at times counterproductive. Improving the system requires a proper diagnosis of the problems, and the political will to create better rules and more accountability.

Lehman Brothers' bankruptcy in September 2008 had major ripple effects throughout the globe and was followed by a deep recession that affected hundreds of millions of people. When housing prices declined starting in 2007, heavily indebted U.S. homeowners began defaulting on their mortgages, exposing the great fragility of the global financial system and the failure of rules that were in place to prevent the excessive buildup of risk.

The mortgage defaults in the runup to the financial crisis were not large relative in magnitude to the global economy. They nevertheless led to a massive global

crisis because of the pervasive use of short-term debt funding by banks and by other financial institutions, the risk these institutions took, and the significant complexity, opacity, and interconnectedness in the system. With little equity funding that could absorb losses on risky assets, many financial institutions became distressed or insolvent, triggering contagion and panic. To prevent a meltdown, central banks and governments provided extraordinary supports to the financial system, particularly to the largest global banks.

With extensive guarantees from the FDIC, trillions of dollars in loans from the Federal Reserve, and hundreds of billions in direct investments by the U.S. government through the Troubled Asset Relief Program (TARP), most U.S. financial institutions did not default and recovered quickly from the crisis (Tooze, 2018). Banks became highly profitable even as mortgage fraud and other wrongdoing led to more than \$300 billion in fines over the last decade. Households, however, continued to suffer from the subsequent recession, exacerbated by heavy mortgage indebtedness and numerous foreclosures (Mian and Sufi, 2014).

A key cause of fragility and inefficiency in the financial system is the excessive use of debt funding by banks and by other institutions. The tax treatment of corporate debt and the various explicit and implicit guarantees banks enjoy perversely encourage and reward reckless risk taking and borrowing. Effective regulation of the funding mix of financial institutions is among the greatest "bargains" in financial regulation, bringing many important benefits at virtually no relevant cost. It will correct market failures directly and at a significantly lower cost than alternative and more complex regulations. Instead, current regulations are

inadequate, and their flawed design creates additional distortions and risks to the system.

Combined with changes in tax and bankruptcy laws, which undermine financial stability, ensuring that banking institutions rely on more equity will bring many benefits, enabling the financial system to serve the economy better without generating as much unnecessary risk and harm. More equity would reduce the intensity of the conflict of interest between bankers and the public and the ability and incentives of the sector, and especially the largest institutions, to grow inefficiently large and to remain opaque, poorly governed, and dangerous.

In this essay, I first discuss the basic economics of corporate funding and why well-designed and effective regulation of banks' funding is highly beneficial. I then point to ways to improve the system and close with remarks that place this policy debate in a broader governance and political context.¹

Corporate Funding and the “Specialness” of Banks

Corporations have many ways to fund their investments. In addition to borrowing, profitable corporations can reinvest their profits or issue new shares of equity. Shareholders absorb losses naturally through reduced value of their shares. If a corporation defaults or has insufficient assets to pay its debts, it becomes insolvent. Insolvency typically leads to a bankruptcy process and, at least, to creditors not being paid in full.

Governments do not usually regulate the funding mix of corporations; most corporations can borrow as much as they want if they can find lenders. The terms of loans are set through negotiations with lenders in private or public markets for corporate debt securities. Corporations may be able to save on their taxes by borrowing, since many governments, including the U.S., consider the interest paid on corporate debt as a deductible expense. Despite this tax advantage, it is rare for healthy corporations outside banking to fund less than 30 percent of their assets by equity, and many thriving corporations borrow little. Retained profits are often the favored source of funding that requires no new issuance of securities to investors.

Heavy borrowing has a dark side.² First, it increases the likelihood of bankruptcy, which depletes the assets through legal costs and disruptions. Second, it intensifies the fundamental conflicts of interest between borrowers (shareholders in the case of a corporation) and lenders regarding investment and funding decisions. The conflicts arise because borrowers benefit fully from the upside of any risk taken while sharing the downside risk with lenders. The decisions made by the managers and shareholders of an indebted corporation may harm creditors and, moreover, they may be inefficient by reducing the combined value of the firm to all investors.

Specifically, once debt is in place, corporate decisions are generally biased in favor of additional borrowing and riskier investments and against reducing indebtedness or making worthy investments that lack sufficient upside. Anticipating the possibility of default, bankruptcy and inefficient investments that harm their interest, creditors protect themselves by requiring higher interest and by placing conditions in the debt contracts. Distressed corporations therefore find it difficult to fund additional investments under favorable terms. Moreover, it is costly to write and enforce debt contracts that prevent excessive borrowing and risk taking, particularly when creditors are fragmented without restricting worthy investments that may benefit all investors.³

The funding mix of banks is starkly different from other companies, consisting almost entirely of debt and very little equity. Even with equity of as little as 3 percent or less relative to their assets (and with problematic measurements that might make the actual indebtedness even heavier), banks seek to make payouts to their shareholders and maintain extremely high indebtedness and borrow to fund more investment.

Because banks have little equity and much of their debt consists of deposits and short-term loans that can be withdrawn quickly, even small losses can cause default or insolvency. If depositors or other lenders become concerned about potential default and lose their trust, they may withdraw their funding, possibly in a panic to ensure they are paid before others. Indeed, concerns about insolvency are a major reason banks can run into “liquidity problems,” suffer from runs and panics and lose their funding, or need support. Having more equity would reduce the risk of banks running into solvency. Pure liquidity problems that arise because assets cannot easily be converted to cash do not usually cause defaults

in banking because central banks stand ready to lend to banks to prevent their defaults.

Deposits and short-term loans benefit from their ability to easily convert to cash. This “liquidity benefit” does not imply, however, that heavy indebtedness and little equity is efficient for banks. To the contrary, as discussed in Admati and Hellwig (2019), the outcomes in *laissez-faire* banking markets, without government intervention of any sort, involve excessive borrowing with inefficiently high likelihood of default that jeopardizes the liquidity benefits of banks’ debt.⁴ Regulation requiring more equity helps create beneficial commitments for banks to avoid the temptation of excessive borrowing that would result under *laissez-faire*.

The need for regulation is even stronger when there is collateral harm to the system or the economy from individual institutions, or many at the same time, becoming distressed or insolvent and possibly defaulting. Guarantees feed and enable the “addiction” to borrowing, associated with heavy corporate indebtedness explored in Admati et al (2018). Since depositors are passive and their claims not backed by collateral, banks will continue to borrow and take risk even if they are insolvent, unless they are stopped by regulators, banks. Moreover, banks have incentives to repeatedly shorten the maturity of their debt to benefit at the expense of existing creditors or taxpayers, which Brunnermeier and Ohemke (2013) refer to as a “maturity rat race.”

For example, banks can use some assets bought with depositors’ funding as collateral for additional borrowing that has higher priority over deposits.⁵ Deposits might be used, in particular, to purchase Mortgage-Backed Securities that are then posted as collateral to obtain more loans on favorable terms using “sale and repurchase agreements” (“repo”) transactions. Such arrangements receive preferential treatment under current bankruptcy laws, making them ever more attractive. Rather than improve outcomes for society, however, these strategies exploit and endanger existing creditors and taxpayers.⁶

The mantra in banking that “equity is expensive” neglects to make the critical distinction between private and social costs. The “costs” of banks using more equity are *entirely private* and incurred by a small set of individuals who would be prevented from benefitting at the expense of others. From society’s perspective, it is

having *too little* equity in banking that is expensive. The “specialness” of banks is therefore that they are allowed to get away with being as inefficient and reckless as they are.⁷

Political economy, confusion and willful blindness are key to understanding why the financial sector is able to maintain its privileges and prevent beneficial regulations. Symbiotic relations between banks and governments create many forms of capture (Admati, 2016). Senator Durbin of Illinois captured the situation by declaring in 2009, just after the financial crisis, that banks “still own the place,” referring to Capitol Hill. Those who benefit from the status quo are able to muddle the debate with misleading narratives.

Highly Beneficial: More Equity, Fewer Debt Subsidies

Prior to the expansion of safety nets for banking (in the form of central banks, deposit insurance, and implicit guarantees), banks maintained much higher equity levels than they have in recent years. As partnerships in the 19th century, for example, equity often accounted for 50 percent of banks’ assets. Since owners were not protected by limited liability, depositors had recourse to the owners’ personal assets if the banks’ assets were insufficient. Equity levels of 20 or 30 percent of total assets were common early in the 20th century, and in the U.S. shareholders had double, triple or unlimited liability until the deposit insurance was established. As safety nets expanded, depositors and other creditors were less concerned and bank shareholders and managers chose to have much less equity and were able to do so without regulations to counter the incentives.

The financial system has become more complex and opaque in recent decades, as well as larger relative to the economy in many developed economies. The growth of securitization and derivatives markets enabled more risk sharing, but it also allowed institutions to take more risks and obscure them from stakeholders.⁸ Regulators ignored risks that built up dangerously, sometimes hiding “off balance sheet” and in the entities in the so-called “shadow banking system.” The growth of the financial sector and of the largest financial institutions has largely been driven by trading within the sector rather than investments in the real economy.⁹ The 2007-2009 financial crisis became the “unintended

consequences” of this massive regulatory failure. Not only were the regulations inadequate, their poor design introduced distortions that increased the fragility of the system and exacerbated the problems. For example, institutions incurred massive losses from investments that regulators had considered perfectly safe.

After the crisis, regulators sought reforms, but they failed to learn the full lessons and proceeded to maintain the overall approach, thus continuing to tolerate a distorted and fragile system. For example, one thorny issue is measurements of indebtedness, particularly in the context of complex derivatives and off-balance-sheet commitments. Accounting-based measures and the use of risk weights in an attempt to calibrate requirements to risk have made regulatory measures quite uninformative for indicating the true strength of any institution.¹⁰

Admati and Hellwig (2013a, Chapter 11) and Admati (2016) summarize the problems with the regulations and propose improvements as well as transition to a better system. Twenty prominent economists (Admati et al, 2010), for example, recommended *at least* fifteen percent, as compared to the 3 percent, in equity relative to total assets required by the 2010 Basel III international accord.¹¹

With more equity, banks would be in a better position to serve the economy even after incurring losses without needing support. They will also be less likely to experience liquidity problems and runs. Moreover, when institutions operate with much more equity funding, any loss in the value of the assets is a smaller fraction of the equity, thus there is less need for distressed asset sales (or so-called “fire sales”). Better yet, by reducing the intensity of the conflict of interest between banks managers and shareholders on one hand, and their lenders and taxpayers on the other, banks with more equity suffer from fewer distortions in lending decisions, including excessive and inefficient risk-taking and underinvestment in some worthy loans.

The easiest way to implement the transition to higher equity requirements is to ban payments to equity until banks are better capitalized, and even requiring that some executive compensation come in the form of new shares rather than cash. It may also be useful for regulators to mandate minimum amounts of new equity issuance each year, with banks that cannot raise equity being viewed as failing a market-based stress test. Any

institution that is too opaque, insolvent, or too big and inefficient to do so should not persist.

Instead of relying on market tests, regulators use so-called stress tests to reassure themselves and the public that the banks are safe enough. These tests set inadequate benchmarks for passing and are based on many strong assumptions. Moreover, they are unable to predict the market dynamics of the interconnected system in an actual crisis, which may come from an unexpected direction. As a result, they give false reassurances.

More equity also provides the easiest and simplest way to reduce the privileges and outsized power of the largest “systemic” institutions, often referred to as “too big to fail.” These institutions are indeed enormously large, complex, and opaque, with assets in the trillions, much larger off-balance sheet exposures, and sprawling operations in many different areas and across the globe.

Vowing to avoid bailouts, the favored approach of regulators and the institutions themselves is to reassure the public that the institutions can “fail” without needing support and causing enormous collateral harm. This approach is flawed in many ways. First, it focuses on treatment of an outbreak in the financial system, when additional equity would act as an obvious preventative measure, reducing the likelihood of failure. Second, the notion that authorities will know just the right moment to trigger a “fail” scenario that would impose losses on creditors, and that the process of doing so would not cause the kind of ripple effect of the Lehman Brothers bankruptcy, is not credible. Indeed, in a crisis when many institutions are failing or near failure, the collateral harm of any process of dealing with the problem would be substantial.

Equity is the simplest, most reliable and most beneficial way to reduce those subsidies while also enhancing the health and safety of the system. Shareholders who are entitled to the upside and who absorb losses without the need to go through complex and costly triggers, are the most obvious candidates to bear the risk.

Suggestions that the largest institutions should be broken up by authorities fail to recognize that the size, complexity and recklessness of these institutions are *symptom* of failed markets and regulations. More equity would be useful because, in addition to reducing the likelihood of costly failure, it is likely, if done properly,

to bring more market pressure from equity investors to cause the largest institutions to break up naturally, similar to how large conglomerates broke up in the 1980s and 1990s. Moreover, as seen in the Savings and Loan crisis of the 1980s and in many other banking crises, the failure of many small banks can cause as much disruption and harm, and may lead to bailouts. Thus, a system with many small but excessively fragile institutions taking similar risks and likely to fail at the same time can present preventable problems.

It is also important to change two sets of counterproductive laws that make the financial system more fragile, and safety regulations in banking harder, by creating a wider gap between what is good for banks and their managers and what is best for society. First, the tax code must be changed to neutralize the advantage of debt over equity funding. Even if banks pay more taxes, this does not represent a cost to society because taxes are used by governments on behalf of the public. The tax effect can also be balanced to have little effect on the taxes banks pay but in a way that does not reward excessive borrowing.¹²

The *Economist* (May 15, 2015) magazine called debt subsidies “a vast distortion in the world economy.” Subsidizing mortgage debt in the tax code makes little sense and has virtually no economic justification. Whereas such subsidies are said to support home ownership, they reward only *borrowing* to buy houses, thus increasing the fragility of households and of the economy to the harm from excessive use of debt (Mian and Sufi, 2015). If home ownership is a policy objective, there are better ways to encourage it, such as providing tax credits towards the down payment (the equity portion) in buying a house.¹³

In addition to the counterproductive tax code that encourages borrowing over equity funding, bankruptcy laws established decades ago, and expanded in 2005, exempt certain repo and derivatives contracts from the normal rules governing creditor behavior in bankruptcy. The expanded “safe harbor” clause was promoted as a way to increase financial stability, but instead it has enabled and encouraged more fragile funding and caused more turbulence during the financial crisis. It further provides special privileges to certain stakeholders, typically other financial institutions, over other lenders. Despite these problems, the counterproductive law has not changed.¹⁴

Large banks also continue to be very opaque.¹⁵ Their recklessness is also evident in the numerous scandals and tens of billions in fines for fraud and other misconduct they routinely pay. Evading rules can go undetected for extended periods, and ultimately leads to relatively small fines viewed as “cost of doing business” and little, if any, personal accountability for executives or the board. These effects breed lawlessness by individuals whose compensation rewards gambling and law evasion, and who rarely pay a personal price when they harm stakeholders and the public. Yet implicit subsidies appear to allow the banking sector as a whole, and particularly the largest institutions, to obtain privileged funding that do not fully reflect the risk they take and to remain profitable despite repeated scandals and fines.¹⁶

Flawed Excuses

The persistent failure to ensure financial stability and the muddled debate about the costs and benefits associated with higher equity are rooted in a mix of confusion and distorted incentives across the individuals in the private sector as well as in government. The situation has prevented engagement on the issues and enabled flawed claims to prevail, starting with an insidious confusion about jargon and continuing with subtly misleading claims or assertions based on inappropriate assumptions. For example, the regulation of banks’ funding mix is referred to as “capital regulation,” but banks are said to “hold” or “set aside” capital, falsely implying that equity funding, which includes funds to be used for making loans and other investment, is akin to idle cash or “a rainy-day fund” that cannot be used for lending. This confusion immediately raises imaginary tradeoffs between lending and equity capital, allowing lobbyists to get away with nonsensical claims (e.g., that increased capital requirements “keep billions out of the economy”). In fact, with more equity banks are better able to make worthy loans at appropriate prices and do so more consistently.

Admati and Hellwig (2015) list 31 distinct flawed claims made in the discussion and provide a brief debunking. Admati (2016, 2017a) describes the actions and the incentives of the many enablers of this situation and thus the dangerous system, including individuals in the private sector, policy, media and academia. Banking scholars are among the enablers when they build models

based on the presumption that markets create efficient outcomes while ignoring critical governance issues and market failures (e.g., due to inability to commit), accepting the system and falsely assuming that ways to change are costly or infeasible.

Among the misleading narratives about financial crises is that they are akin to natural disasters and thus unpreventable. This narrative directs discussion to disaster preparation, akin to sending ambulances to the scene of an accident, rather than to prevention. Enablers also misleadingly use the past failure of regulation that resulted in the growth of the so-called shadow banking system as an argument against regulations. Worse, where simple and cost-effective regulations can help counter distorted incentives, regulators have instead devised extremely complex regulations some of which may not bring enough benefit to justify the costs but which allow the pretense of action.

Enablers often invoke the claim that we must maintain a “level playing field” in regulations and ensure the global competitiveness and success of “our” national institutions. The “success” of banks in Ireland and Iceland before the financial crisis, however, came at an enormous cost to their taxpayers. Just as we should not allow pollution even if another nation foolishly tolerates it, subsidizing recklessness in banking to help them succeed while endangering our citizens and others is bad policy. Moreover, financial institutions compete with other sectors in the economy, including for scarce human capital. Their ability to attract bright individuals whose talents might be better used elsewhere creates additional and often invisible market distortions and harm, likely exacerbating inequality.

Concluding Remarks

The financial Laws and regulations should be designed to reduce the conflict between individuals in the financial sector and what is good for society more broadly. Despite the efforts of some politicians, regulators, public-interest groups, commentators, and academics, new regulations do not fully reflect the lessons of the 2007-2009 crisis. A financial system meant to allocate risk and resources efficiently instead continues to distort the economy and endanger the public. Confusion and the politics of banking regulations remain obstacles to change.¹⁷

Politicians tend to see financial institutions as a source of funding for their favored causes, including political campaigns or other projects that appear to appeal to voters. Turning a blind eye to risk in banking is convenient. Implicit guarantees appear free, and policymakers who tolerate recklessness in banking rarely face political consequences. The public may be confused by the many flawed claims made by the industry and its many enablers, and falls prey to short-sighted promises of cheap credit. Borrowing too much can cause great harm, particularly for the lowest-income households, yet lenders’ own recklessness is tolerated.

In summary, despite a massive financial crisis and regulatory reform, the financial system remains too fragile. Powerful individuals benefit from the fragility, and from the excessive complexity of the regulation, and they get away with maintaining it. Change will not come easily given the entrenched interests of those involved, and the inertia of the system. Appropriate public understanding of the root cause of the problems beyond awareness of some of the obvious symptoms, such as the persistence of too-big-to-fail institutions and many misconduct scandals, and of the true tradeoffs of different policy choices is essential.

The financial sector is an extreme example of deep and broad problems in the nexus of corporate governance and political economy. Corporations claiming to maximize “shareholder value” often cause preventable harm when governments fail to act in the public interest.¹⁸

Anat R. Admati is the George G.C. Parker Professor of Finance and Economics at Stanford University. Contact: admati@stanford.edu

Endnotes

- * This essay was prepared as part of the Economics for Inclusive Prosperity (EfIP) series of policy briefs. Anat Admati, is from the Graduate School of Business, Stanford University. I am grateful to Dani Rodrik and Graham Steele for helpful comments.
- 1 Numerous pieces and other materials on these topics are linked from this website on excessive leverage and risk in banking, <https://www.gsb.stanford.edu/faculty-research/excessive-leverage> as well as from my personal website <https://admati.people.stanford.edu/>
 - 2 See Admati and Hellwig (2013a, Chapter 3).
 - 3 Ensuring that managers would not pass up worthy projects that would have benefitted creditors and increase the total value of the corporation is also extremely challenging to do through debt covenants.
 - 4 Admati and Hellwig (2019) builds on Admati et al (2018), which shows how the conflicts of interests and inability to fully commit cause heavy borrowing to become “addictive” and why the insights are particularly relevant in banking.
 - 5 Insured depositors are so passive that they, and the banks, themselves forget that they are actually creditors and that deposits are part of the banks’ debts. For example, John Stumpf, past CEO of Wells Fargo Bank, made the nonsensical claim that because his bank has a lot of retail deposits, it does not have a lot of debt, and he was quoted in a later story whose title referred to Wells Fargo Bank as “debt averse” saying “the last thing I need is debt.” (The first quote is from “Wells Chief warns Fed over Debt proposal,” Tom Braithwaite, *Financial Times*, June 2, 2013, the second is from “Fed’s Disaster Plan Is Bitter Pill for Debt-Averse Wells Fargo,” Jesse Hamilton and Ian Katz, Bloomberg News, October 29, 2015.) Of course, a truly debt averse Wells Fargo Bank could reduce its indebtedness by retaining its profits or selling new shares. These statements illustrate that despite their extreme indebtedness, banks do not experience the burdens and the market forces that affect other corporations.
 - 6 A repo transaction is economically equivalent to secured borrowing, i.e., borrowing with the use of collateral, but a repo consists of a simultaneous “sale” of the collateral to the lender and a commitment to buy or repurchase it at a future point of time at a fixed price. Under safe harbor provisions in the U.S. bankruptcy code, many financial sector repo lenders can possess the collateral asset even if the borrower goes into bankruptcy that would typically freeze debt claims.
 - 7 For a discussion of implicit guarantees and some of the efforts to estimate them, see Admati (2014) and Gudmundsson (2016).
 - 8 See Admati and Hellwig (2013, Chapter 5) and Eisinger and Partnoy (2013).
 - 9 See, for example, Haldane et al. (2010) and Turner (2010).
 - 10 Singh and Alam (2018) show that current measures of indebtedness are misleading because they do not account properly for exposures off balance sheet. The authors assess these exposures to be larger than in 2007 just ahead of the financial crisis.
 - 11 Cochrane (2013) captures the spirit of the answer, namely requiring enough equity that it no longer matters because the downside risk is borne by shareholders.
 - 12 Roe and Troege (2018) discuss the distortion created by tax subsidies of bank debt and propose changes specific to banking.
 - 13 Jorda et al. (2016) show that banks and households have become heavily indebted through mortgages in the second half of the 20th century and that mortgage credit has been important in understanding the increased financialization, the fragility of advanced economies, and the dynamics of business cycles.
 - 14 Morrison et al. (2014) argue convincingly that the safe harbor rules for repos should apply much more narrowly and the 2005 law should be repealed and the 1984 version remain in effect. See also Partnoy and Skeel (2007) and Jackson and Skeel (2012), which also describe the similar bankruptcy treatment of derivatives.
 - 15 Eisinger and Partnoy (2013), which examined the financial statements of Wells Fargo Bank, quotes many investors and accounting experts stating that the large banks are “uninvestible.” See also Singh, Manmohan Alam (2018).
 - 16 See Admati and Hellwig (2013a, chapters 8, 9 and 13) and Admati (2014) on the incentives for recklessness. A CNBC headline in March, 2017 captures the notion that fines are “cost of doing business” announcing: “Banks Have Paid \$321 Billion in Fines Since the Financial Crisis (But They have Made Nearly \$1 Trillion)” (<https://www.cnn.com/2017/03/03/banks-have-paid-321-billion-in-fines-since-the-crisis.html>)
 - 17 Admati (2017a) cites some of the terms in social psychology that apply to the various blind spots in this area, such as willful blindness, collective moral disengagement. See also (Jost 2017) on system justification. Pfleiderer (2018) discusses the misuse of models in economics and finance. Many materials at various lengths, including videos and slide presentations with visuals, are available at <https://admati.people.stanford.edu/advocacy>
 - 18 See Admati (2017b).

References

- Admati, Anat R., Franklin Allen, Richard Brealey, Michael Brennan, Markus K. Brunnermeier, Arnoud Boot, John H. Cochrane, Peter M. DeMarzo, Eugene F. Fama, Michael Fishman, Charles Goodhart, Martin F. Hellwig, Hayne Leland, Stewart C. Myers, Paul Pfleiderer, Jean Charles Rochet, Stephen A. Ross, William F. Sharpe, Chester S. Spatt, and Anjan Thakor (2010), “Healthy Banking System is the Goal, Not Profitable Banks,” *Financial Times*, Nov. 9.
- Admati, Anat R. 2014. “Examining the GAO Report on Expectations of Government Support for Bank Holding Companies,” Testimony for hearing of Senate Committee on Banking, Housing and Urban Affairs Subcommittee on Financial Institutions and Consumer Protection, July 31, 2014.
- . 2016. “The Missed Opportunity and Challenge of Capital Regulation.” *National Institute Economic Review*, February, R4-R14.
- . 2017a. “It Takes a Village to Maintain a Dangerous Financial System,” *Just Financial Markets? Finance in a Just Society*, Lisa Herzog, editor, Oxford University Press.
- . 2017b. “A Skeptical View of Financialized Corporate Governance,” *Journal of Economic Perspectives* 31(3), 131-150.
- Admati, Anat R., Peter M. DeMarzo, Martin F. Hellwig, and Paul Pfleiderer. 2013. “Fallacies, Irrelevant Facts, and Myths in the Discussion of Capital Regulation: Why Bank Equity Is Not Socially Expensive.” Working Paper (original draft August, 2010).
- . 2018. “The Leverage Ratchet Effect.” *Journal of Finance* 73 (1), 146 – 198.
- Admati, Anat R. and Martin F. Hellwig. 2013a. *The Bankers’ New Clothes: What’s Wrong with Banking and What to Do About It*. Princeton University Press.
- . “Does Debt Discipline Bankers? An Academic Myth about Bank Indebtedness.” 2013b. Working paper.
- . 2015. “The Parade of Bankers’ New Clothes Continues: 31 Flawed Claims Debunked.” Working paper.
- . 2019. “Bank Leverage, Welfare and Regulation.” in *Systemic Risk in the Financial Sector: Ten Years After the Great Crash*, Douglas W. Arner, Emiliós Avgouleas, Danny Busch and Steven L. Schwarcz, editors, CIGI Press, forthcoming.
- Brunnermeier, Markus, and Martin Oehmke. 2013. “The Maturity Rat Race,” *Journal of Finance*, 68:483-521.
- Cochrane, John H. 2013. “Running on Empty.” *Wall Street Journal*, March 1.
- Gudmundsson, Tryggvi. 2016. “Whose Credit Line is it Anyway: An Update on Banks’ Implicit Subsidies.” IMF Working Paper 16/24.
- Eisinger, Jesse and Frank Partnoy. 2013. “What’s Inside America’s Banks?” *The Atlantic*, January-February.
- Haldane, Andrew, Simon Brennan, and Vasileios Madouros. 2010. “What is the Contributions of the Financial Sector: Miracle or Mirage?” *The Future of Finance, the LSE Report*.
- Jordà, Òscar, Moritz Schularick and Alan M. Taylor. 2016. “The great mortgaging: housing finance, crises and business cycles” *Economic Policy*, 31 (85), 107-152.
- Jost, John T. 2017. “A Theory of System Justification.” Psychological Science Agenda Science Brief.
- Mian, Atif, and Amir Sufi. 2015. *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again*, University of Chicago Press, Chicago, IL.
- Morrison, Edward R., Mark J. Roe, and Christopher S. Sontchi (2014). “Rolling Back the Repo Safe Harbors,” *The Business Lawyer*, 69: 1015-1047.
- Partnoy, Frank, and David Skeel. 2007. “The Promise and Perils of Credit Derivatives.” *University of Cincinnati Law Review*, 75. 1019-1051
- Pfleiderer, Paul. 2018. “Chameleons: The Misuse of Theoretical Models in Finance and Economics.” *Economica*, Forthcoming.
- Roe, Mark J. Michael Troege. 2018. “Containing Systemic Risk by Taxing Banks Properly.” *Yale Journal on Regulation* 35 (1), 181-231.

Singh, Manmohan, and Zohair Alam, 2018. “Leverage – A Broader View”, IMF Working Paper 18/62, International Monetary Fund, Washington, D.C.

Turner, Adair. 2010. “What Do Banks Do? Why Do Credit Booms and Busts Occur? What Can Public Policy Do About It?” *The Future of Finance, the LSE Report*.