CHAPTER 6: THE DATA REVOLUTION AND THE TRANSFORMATION OF SOCIAL PROTECTION

Torben Iversen

Philipp Rehm

[Fortcoming In Marius R. Busemeyer, Achim Kemmerling, Paul Marx, and Kees van Kersbergen (eds)*, Digitalization and the Welfare State.* Oxford University Press 2021.]

Abstract

A central function of the welfare state is to provide social insurance. Most scholarship assumes that social insurance cannot be provided effectively through the market, mainly due to incomplete and asymmetric information. But, while this assumption may have held in the past, the data revolution is making it untenable today. This chapter asks what happens to the politics of social protection, and to inequality, when information about risks to health, life, employment, credit, and so on, becomes more widely available and shareable. It offers a new integrated analytical framework to help us better understand the consequences of big data for social policy and inequality. The developments in the private insurance market—fueled by the information revolution—can be taken as a harbinger of what is to come in social insurance. The information revolution potentially undermines majority support for public social policy programs, might also lead to widespread privatization, and it could even end the welfare state as we know it. Our argument implies that as more information can be credibly shared with insurers, private markets for social insurance are becoming more feasible. However, pressure to marketize is not the same as privatization. Bearing this in mind, this chapter discusses the main factors that will prevent private insurance, even in the presence of extensive and symmetric information. The main conclusion is that the information revolution will lead to more polarization of welfare state politics. Moreover, at least in some areas of social insurance, it might also lead to the (gradual) privatization of risk pools that are currently socialized. Either way, solidaristic risk-sharing will come under political attack, and inequality in social protection and incomes is likely to increase.

Introduction

There is a broad consensus in the political economy literature that a central function of the welfare state is to provide social insurance (Baldwin 1990; Esping-Andersen 1990; Iversen/Soskice 2001; Moene/Wallerstein 2001; de Swaan 1988). Underlying most of these analyses is an assumption, often implicit but virtually universal, that social insurance cannot be provided effectively through the market, mainly due to incomplete and asymmetric information (Akerlof 1970; Barr 2012; Boadway/Keen 2000; Stiglitz 1982). But, while this assumption may have applied in the past, the data revolution is making it untenable today. This chapter asks what happens to the politics of social protection and to inequality when information about risks to health, life, employment, credit, and so on, becomes more widely available and shareable.

A hint of what is to come can be gleaned from the life insurance market, where ICT is radically transforming the status quo. For example, John Hancock Life Insurance, a major player in the US American market, introduced a policy that calculates annual premiums partially based on data collected by an "activity tracker" that policyholders receive for free when they sign up. These types of devices can track and instantly share (via an app) things like: steps and stairs taken, active minutes, calories burned, heart rate, sleep quality and blood pressure, among others. The company marketed this life insurance policy as "an innovative solution that rewards you for living a healthy life. In fact, the healthier you are, the more you can save."[1]

[1] FAQs on the John Hancock Vitality Program (https://www.johnhancockinsurance.com/vitality-program/vitality-faq.html, retrieved March 11, 2021). President and General Manager of John Hancock Insurance, Michael Doughty, assures customers: "You do not have to send us any data you are not comfortable with," though he points out: "The trade-off is you won't get points for that." (New York Times, April 8, 2015, p. B1).

And it is not just insurance companies getting in on the action. The leading technology companies—Apple, Alphabet, Amazon, Microsoft, etc.—are all committing huge resources to develop a new data-based health industry, where, *inter alia*, doctors can interact with AI-enabled databases, and individuals can easily share their information with insurance companies. Similar efforts are underway in credit markets, where detailed information about credit history is linked to a trove of data on income, occupation, residence, credit history, etc.

There is currently no integrated analytical framework to help us better understand the consequences of big data for social policy and inequality. Our chapter seeks to provide such a framework and apply it to the political economy of social protection, with an emphasis on the knowledge economy and due consideration to the role of national political and regulatory institutions.

Theory

Akerlof's (1970) seminal *QJE* article "On the Market for Lemons" presents a key reason for the breakdown of markets: asymmetric information and the associated problem of adverse selection. Rothschild and Stiglitz (1976) and Stiglitz (1982) were the first to model the logic for insurance markets, with more recent extensions summarized in Boadway and Keen (2000), Przeworski (2003), and Barr (2012). In these models, individuals know their risk types, but insurers do not.

Although the technical details are complex, the logic of the model is as simple as it is compelling. If insurers do not have individual information about risks, they will only know the mean risk in the pool of the insured, which can be inferred from insurance payouts. Based on this mean, the insurer can set a flat premium that is just high enough to settle all (actual and expected) insurance claims. If people have information about their own risks, such an insurance plan will be attractive to everyone whose risks are above the mean; they pay in less than they expect to collect. But the opposite is true for those whose risks are below the mean, because they will end up subsidizing high-risk types. Depending on the degree of risk aversion, some individuals who are not too far below the mean will still buy insurance, but others at low risk will opt out entirely. When low-risk types opt out, the result is that the average risk in the insurance pool rises, which, in turn, will prompt others to opt out, etc. This is the adverse selection problem, which led Akerlof to conclude that there would only be a market for “lemons”, although the process may in fact stop short of complete market failure because people with high risk aversion and moderate risks may stay in the pool despite the fact that they are subsidizing bad risks. In any event, there is market failure in the sense that many who want insurance will end up without insurance or with less insurance than they would like. This underinsured group may include both low-risk types, who opt out, and high-risk types with low incomes who cannot afford the higher premiums as good risks leave.

The economic analysis usually ends here, and public provision is “explained” as a solution to a problem of private under-provision—i.e. market failure. But, politically, efficiency cannot be what drives the introduction of a public system. Historically, public systems were constructed as single pools with a common premium paid through income taxes (or similar income-related contributions), and in such all-encompassing pools, there must be some low-risk types who pay “too much” for their insurance and would want to opt out, just as there is with private insurance. The difference between a public and a private system is that the state can compel people to pay into the system; opt-out is not an option. From this perspective, the welfare state emerged not for reasons of efficiency, but through coercion of the majority.

The transition to a public system therefore comes about, assuming democratic decision-making prevails, only because a majority prefers public to private provision. This is not a high bar to cross because in the private market adverse selection usually ensures that those with moderate risk and income will be saddled with very high insurance rates so as to pay for themselves and for those with higher risks, whereas good risks will not pay into the system at all. In the public system, those with median risks will still subsidize those at higher risk, but they will also be subsidized themselves by those with lower risks (and typically higher income). So, asymmetric information creates a political path to the welfare state. This is summarized as scenario (2) in Table 1.

This account implies that the rise of the welfare state is driven by the middle class, but it also suggests that it is a contentious process. In this sense, it is entirely consistent with "power resource theory" (Esping-Andersen 1985; Korpi 1983; Stephens 1979) because income and risk determine how much an individual contributes and benefits, and income and risk are, in turn, correlated with class. Once placed in a public system, the logic translates into fiscal preferences: those with high income and low risk want to spend less on public insurance than those with low income and high risk. Much of the literature on the political economy of public opinion is concerned with documenting these differences in preferences, and how political parties respond to them.

An important qualification, however, is that the degree to which people are divided over social spending depends on how well-informed they are about their own risks. If people are uncertain about where they are in the risk distribution, their prior is close to the mean risk, and support for spending will converge to the mean level.[2] As a consequence, the greater uncertainty, the less polarization in preferences. A very simple way of expressing this general insight is that *class conflict increases with information*. This echoes Rehm's (2016) observation that homogeneity in the risk distribution reduces conflict over the welfare state, but here homogeneity is *induced* by lack of information—the distribution of actual risk does not change. This is illustrated by scenario (1) in Table 1. Of course, some will be well informed about their risks, but in the early phase of welfare state development it is reasonable to assume that the vast majority only had limited information. In such a world of uncertainty, as long as a public system forces at least some (known) low-risk types into the insurance pool, the welfare state is likely to enjoy widespread cross-class support. We may see this as the foundation for the "Golden Age" of welfare state expansion.

Table 1: Information and social insurance

		Distribution of information	
		Asymmetric	**Symmetric**
Level of information	**Low**	(1) High uncertainty case (Consensus)	
	High	(2) Classic adverse selection case (Polarization)	(3) Efficient market case (for majority able to pay)

[2] If *i* receives signals about his or her true risk, p_i, from a noisy environment in which the overall mean is $\overline{p}$ then $p_i^o = \alpha \cdot p_i^s + (1-\alpha) \cdot \overline{\overline{p}}$, where p_i^s is a signal drawn from a distribution that is centered on the individual's true risk (p_i) and α is a measure of the "precision" of that signal, which in our model, equals the private information available to *i* (a formal proof can be found in Iversen and Soskice (2015), Appendix B.). In the extreme case of no information ($\alpha = 0$), $p_i^o = \overline{\overline{p}}$, so the range is zero; at the other extreme of complete information, $p_i^o = p_i$, the range equals the difference between those with the lowest and those with the highest risk.

Scenario (3) in Table 1 is one in which both insurers and private individuals have complete information. Akerlof and Stiglitz did not discuss this case, since they were interested in exploring the consequences of private information. However, the symmetric information case is important to our story. Even when privacy protection limits the ability of insurers to acquire individual information, as is the case with medical records, for example, people may *choose* to share that information. The example in the introduction of using monitoring devices to reduce insurance premiums provides the intuition: good risks pay lower premiums if they can share their type.

This logic applies to the important area of private health data where the level and credibility of information has vastly improved over time. There are three related forces driving this trend. First, the general advance of medicine has increased the detail and reliability of diagnostics (Shojania et al. 2003). Second, the explosion in the number and variety of tests that can be done by certified laboratories has made it possible to share this information credibly; DNA diagnostics in particular promises to offer an order of magnitude more information about health risks than in the past. Finally, computing power combined with AI has made it possible to classify individuals in risk groups much more accurately than in the past.

The fact that individual information can be acquired by, and credibly shared with, would-be insurers mitigates the asymmetric information problem and opens up the possibility of insurance being provided efficiently through the market to those who are able to pay. For each group of individuals with identical risk profiles, there would be a separate insurance pool/plan with its own cost and replacement rate. Realistically, there is likely to be some modest risk heterogeneity within groups that is unknown to the insurer and therefore pooled. As long as insurers have enough information to distinguish members of different groups, we would get a series of distinct insurance pools and plans.

In this brave new world of near-complete information, there would be an effectively functioning market for the "creampuffs"—people with low risks that insurance companies crave. In fact, anyone with risks below the mean would be better off in such a world, assuming that private provision is no less efficient than public provision.[3] Again, this is because, in a public system, all those with below-average risk subsidize those with above-average risk. Another implication is that those with the highest risk, who also tend to have the lowest incomes, may be unable to afford private insurance. For example, low-income people at serious risk of diabetes may be unable to effectively insure against that risk if is known to insurers. In this sense, the market is obviously not efficient, but it may work efficiently for a majority.

Of course, high-risk types would want to protect their privacy. Some individuals with lower risks may also want to do so on principle, but this will be costly because "refusers" are automatically placed in the same pool as high-risk types, driving up their premiums. Everyone in the pool with risks below the group average will therefore have a financial incentive to divulge their information to get a cheaper plan. But if they do share their information, the same will be true for those with risks below the mean in the remaining pool, and so on. This process will continue until all the "lemons" have been called out. This is Akerlof's logic *in reverse* because the result will now be a segmented private market for risk where all information is common knowledge, and with an uninsured group of high risks who cannot pay. It is clear from this analysis that privacy laws are *not* sufficient to remedy the problem of information that can be credibly shared.

[3] This is of course an issue of considerable debate. Suffice it to say here that the assumption helps to home in on the effect of information on the direction of change in distributive politics.

Would there be majority support for privatization? Based on our assumptions, the answer may be yes if the risk distribution is bottom-heavy. This is true in the case of health risks,[4] and it is ordinarily also true in the case of unemployment risks (Rehm 2016). In an "up or down" vote, self-interested voters may therefore support privatization. Again, this conclusion only holds in a world of complete and shared information, and there are many complicating factors, most notably the relative cost of private insurance compared to a single-payer system. Our main argument is that big data will make the public system increasingly contested, and it could give way to private alternatives or market-conforming reforms of public systems.

Broadly speaking, our argument implies that the information revolution results in pressures toward more market-conforming provision. Pressure to marketize is not the same as privatization, however. First, markets can be preempted by allowing greater differentiation in the public sector. In this case, we can think of private markets as having an effect through "shadow prices": the public sector mimics the private in terms of choice and prices. This would imply policies that allow more differentiation and choice in the public system, use of credits for supplementary private insurance, cuts in benefits for high-risk groups (such as refusal to cover procedures for obesity), and introducing high copayments (which are highly regressive).

Second, pressures for marketization will be tempered by governments using measures such as "price non-discrimination" clauses to rule out private markets. These prevent insurers from offering better deals to those with lower-risk profiles. Since parties on the left of the political spectrum tend to represent

[4] While there is no direct data on risk, health spending is highly concentrated. In 2009, about half of US healthcare spending went to just five percent of the population (National Institute for Health Care Management 2012). Of course, much of this spending is on the elderly, and everyone grows old, so people must worry about insurance as they age. We consider this issue below. During pandemics, the health risk distribution may be top-heavy.

constituencies that are lower income and higher risk, left governments are more likely to promote such policies. Uncertainty about the cost and effectiveness of private alternatives can also deter voters from supporting privatization, and the exchange of certain benefits for uncertain gains is known to cause resistance. When markets are blocked for any reason, increased information instead causes polarization in preferences regarding the level of public provision and the distribution of the costs. This polarization can intensify in the presence of effective private options because such options make the feasibility of markets clear—even if it is hard for individuals to take advantage of these options because of the double payment problem (paying for private plans, while also being taxed for public ones).

Our argument implies the following two hypotheses, which we will explore in the next section:

> H1: As information and credible sharing of information improves, private insurance markets will rely on such information, when it becomes available, for purposes of risk classification and premium calculation.

> H2a: Where private insurance markets exist, they expand as information and credible sharing of information improve.

> H2b: The expansion of private insurance markets will be more rapid under right than left governments.

Empirics

In this section, we explore two implications of our theoretical framework. To start with, we illustrate the basic logic of our argument by reporting some trends in private insurance markets that are the result of increased information, and the emergence of the technology to credibly share it. We then test the hypothesis that more information will lead to an expansion of insurance, using the life insurance market as an example.

Individualization of Life and Health Insurance

Car insurance is a good example of how information technologies can radically transform an insurance market. A key problem in this market is asymmetric information: driving behavior is a major factor in accident proneness and should therefore be used in premium calculations, but insurance companies have traditionally been unable to monitor it. Advances in information technology have alleviated this asymmetric information problem, which radically transforms the status quo.

In particular, insurance companies can now use tracking devices that collect—and instantly transmit—data on driving behavior, such as distance driven, acceleration, braking events, cornering forces, speeds relative to speed limits, and so on. This enables insurance companies to offer individualized insurance premiums that are directly tied to observed driving behavior, and there are now more than a dozen companies in the US offering these “pay as/how you drive” policies. These plans are attractive for safe drivers which insurance companies hope to sign up.

Similar developments can be observed in the life and health insurance markets. In these domains, tracking devices are one tool for credible information sharing, with genetic testing another. Information obtained through these channels allows companies to offer individually targeted insurance plans, based on detailed risk classification. In this business model, insurance companies worldwide team up with firms like Discovery Limited which develops wellness programs branded as “Vitality – A wellness solution that changes the way insurance works”. Vitality uses data on consumer behavior, which are collected by fitness trackers (such as Fitbit, Jawbone, Misfit, Apple Watch) and transmitted to the company or insurer. Additional information, such as purchasing data, is sometimes collected as well. This detailed, constant, and instant tracking of consumers is useful for health and life insurance companies alike, as explained on the company’s (now defunct) website:

> “Insurers traditionally use risk rating factors to access and underwrite risk. These include age, gender, socio-economic status as well as smoker status and medical history. These risk factors are mostly static and offer a limited view of a person’s risk. A person’s health behavior, however, provides a more accurate risk indication. Vitality, with its 17 years of wellness experience, data and understanding of wellness behavior, adds an additional dynamic underwriting rating factor. It takes into account the impact of chronic diseases and lifestyle factors, such as smoking, level of exercise, diet, alcohol consumption, blood pressure and cholesterol on a person’s risk profile. By integrating Vitality with insurance products, we have developed a scientific and dynamic underwriting model that uses high-quality data about a person’s health, wellness, credit card spending and driving behavior to assess their risk more accurately over time. This results in: better benefits, lower and more accurate risk pricing, better selection, lower laps rates, [and] better mortality and morbidity experience.”[5]

Tracking devices (“wearables”) are increasingly common in the life insurance market. And health insurance coverage plans tied to these tracking devices—frequently in combination with workplace

[5] https://web.archive.org/web/20160322151205/http://www.vitalitygameon.com/vitalitygameon/ (retrieved March 11, 2021).

wellness programs, which often perform health risk assessments and biometric screenings—are gradually being rolled out as well. At one point, policyholders with Oscar Health Insurance in the US, for example, received a free step tracker and could earn up to US$1/day for taking a particular number of steps. In a similar vein, health insurance company UnitedHealth and chipmaker Qualcomm have teamed up to develop a wearable device tied to a coverage plan that incentivizes health behavior by paying up to US$4/day to a covered employee and their spouse if they reach certain targets.[6]

The new tracking systems have become most widespread on the life insurance market and include John Hancock Life Insurance (US), Prudential's Vitality Health (UK), AIA Australia life insurance and MLC On Track (Australia), and Generali (Austria, France, Germany). But they are equally useful for health insurers, and currently more than 100,000 employees at an undisclosed number of employers in a dozen US states are using wearables through UnitedHealth plans.

Yet, this is surely only the beginning. Big Tech is committing huge resources to the advancement of a new data-based health industry, using a variety of related strategies. Apple and Alphabet are developing new tracking technologies—for heart rhythms, sleep patterns, circulation, and even blood sugar levels—using wearables (including a smart contact lens), and microchips can now be implanted directly into the body to constantly monitor health and relay information. Alphabet has recently created a new research unit, called Verily Life Sciences, to develop these technologies using AI-based approaches to data analysis, and Microsoft's Healthcare NeXT is focused on collecting huge amounts of individual data from a variety of

[6] Technological progress makes tracking devices ever more sophisticated. One example is the Kolibree toothbrush whose 3-axis accelerometer, 3-axis gyrometer, and 3-axis magnetometer can decipher detailed subtle movements in order to provide real-time feedback that gets transferred to the brusher's smartphone via Bluetooth, and from there can be shared with a dentist. Of course, it could also be shared with a dental insurance company.

sources and transferring it to cloud-based systems, including a virtual assistant that takes notes at patient-doctor meetings using speech recognition technologies (Singer 2017).

As we will document below, the number of reliable tests that can be conducted by independent laboratories has also greatly increased over time, and these data can be combined with data from the new tracking algorithms to produce detailed profiles of individual health parameters with enormous predictive power. The promise of "personalized medicine" is based on such individual information, and former US President Obama's Precision Medicine Initiative reads like an impassioned call for more data on people's underlying health risks—"including their genome sequence, microbiome composition, health history, lifestyle, and diet." As AI crunches these numbers, much greater risk differentiation in insurance policies becomes feasible, which in turn expands the reach of markets, which is ultimately likely to result in greater inequality in coverage and cost. An extreme scenario would be if advances in technology and medical knowledge were to enable companies to predict someone's medical future with great accuracy, rendering bad risks uninsurable in private markets, while offering good risks a whole range of attractive options.

Information and Private Market Penetration: Life Insurance

To explore the effect of information on private market penetration, we turn to information about health and the development of life insurance markets. Apart from modest programs for survivor's (widow's) pensions, the public system offers no life insurance. This is therefore an obvious area of potential private expansion as more medical information becomes available that can be credibly shared.

The principle of life insurance is very simple: people pay a predetermined monthly premium as long as they live, and the insurance company pays a predetermined amount to survivors when the policyholder dies. If the policyholder dies before the cost of the payout (adjusting for interest) is covered, the insurer loses, while the family of the insured gains. For the insurer to calculate the insurance premium it is therefore essential to be able to calculate life expectancy of potential policyholders accurately. The adverse selection problem is obvious in this context because people will buy plans that assume they will live longer than they themselves expect to live. The first life insurance plans were restricted to pools where members shared so many traits that their life expectancy could be calculated with great accuracy. The Scottish Presbyterian Widows Fund, commonly credited as the world's first modern life insurance scheme, was restricted to Scottish Presbyterian clergymen—a very homogenous group with high entry barriers and an average life expectancy that was easy to calculate from carefully-kept church records.

Modern life insurance schemes rely on information about individual health, complemented by demographic and related data. The expectation is that better information regarding health risks leads to larger life insurance markets. Our *dependent variable* is life insurance market penetration measured as a ratio of direct gross life insurance premiums to gross domestic product (GDP). This measure was developed by the OECD and "represents the relative importance of the [life] insurance industry in the domestic economy" (OECD 2015). We have data covering 22 advanced economies for the period around 1983–2013. (The Appendix provides more details on the data, as well as methods, variables, and results).

The key *independent variable* is private information that can be credibly shared with insurers. We do not, of course, have direct access to private information, but such information is reflected to some extent in the availability of diagnostic tests. Accurate tests by independent laboratories are one element of what insurance companies need to distinguish risk groups, and such tests—based on blood, saliva, urine, tissue, and increasingly also genetic samples, as well as CT and MRI scanning—have become much more common, accurate, and affordable. A striking example is the cost of sequencing the human genome, which has dropped from about US$300 million in 2001, US$1,000 in 2014, to less than US$200 in 2018.[7] Correspondingly, the number of personalized gene-based diagnostic tests and treatments available in the US has also risen from 13 in 2006 to 113 in 2014 (Personalized Medicine Coalition, 2014). The number of standard tests which can be carried out on a simple penetration blood sample increased from 130 in 1992 to 319 in 2014 (Pagana/Pagana 1992; Pagana et al. 2014).

We can also trace the development of diagnostic capabilities via an authoritative and widely used indexing system for diagnostic tests operated by the National Library of Medicine. It maintains a list of 27,000 or so “Medical Subject Headings [MeSH]” (Coletti/Bleich 2001) that are designed to map the entire biomedical field based on English-language academic journals. The MeSH classification includes a hierarchical tree structure where one sub-branch indexes terms related to “Diagnosis” (E01). In 1971, there were 277 index entries; there were 450 in 1981, 600 in 1991, 701 in 2001, 914 in 2011, and 1,067 in 2014.

Tests can be used to predict life expectancy by disease, and the more tests are conducted, the greater the accuracy of these predictions. The World Health Organization (WHO) collects detailed data on mortality by cause, and calculates the “Potential Years of Life Lost” (PYLL) for each major cause of death (cancer, cardiovascular deceases, AIDS, etc.). PYLL is the absolute difference between how long people diagnosed with a particular disease actually live and the average life expectancy (weighting deaths occurring at

[7] According to The Economist (March 14-20, 2020, p. 5), “the first genome cost, by some estimates, $3bn.” Moreover, the costs for sequencing a human genome have been falling faster than Moore’s law (The Economist, March 14-20, 2020, p. 8).

younger ages more heavily).[8] PYLL has become more detailed and accurate over time, it varies by country and year, and it is available for a broad set of countries.

The business of life insurance is to predict life expectancy, and PYLL is precisely the information they need to estimate expected payouts for people with particular conditions. For healthy buyers, insurers will have to rely on diagnostic information that is predictive of such conditions, and we do not have access to this information as researchers. Yet, PYLL carries useful indirect information about risks. This is because accurate and timely diagnosis is a necessary condition for effective treatment, and therefore for a lower PYLL. For example, Hereditary Amyloidosis is a condition that is caused by an inherited genetic mutation, which can be identified through DNA testing long before symptoms arise.[9] Once symptoms appear, there are blood and tissue tests that can pinpoint the exact form of the disease, which in turn decides treatment. Most who are diagnosed with Hereditary Amyloidosis eventually die from heart or kidney failure, but early detection and treatment – ranging from a strict diet to drugs and even liver transplants – create a wide PYLL range. Needless to say, a late or inaccurate diagnosis increases PYLL.

In general, better diagnosis should be negatively related to PYLL, and this is indeed what we find when we regress the MeSH data on the PYLL series, along with a set of control variables.[10] Better diagnostics leads to better treatment, which reduces premature death. Hence, PYLL is also a good indicator of underlying

[8] A low PYLL is, however, not a necessary condition for the availability of information because some diseases, especially soon after they are discovered, are not treatable even if they can be accurately diagnosed (AIDS was a case in point).
[9] A low-cost DNA sequencing service such as 23&Me tests for a common variant.
[10] Controls are health insurance coverage (% of population), total health expenditures (% of GDP), and economic growth rate.

risks that are not directly observed as a disorder.[11] Specifically, if countries where people die earlier from particular diseases have fewer diagnostic tests available then this will limit the scope for the life insurance industry to flourish because it depends entirely on an established infrastructure of laboratories, testing technology, and expertise to estimate individual life expectancy and therefore limit adverse selection.

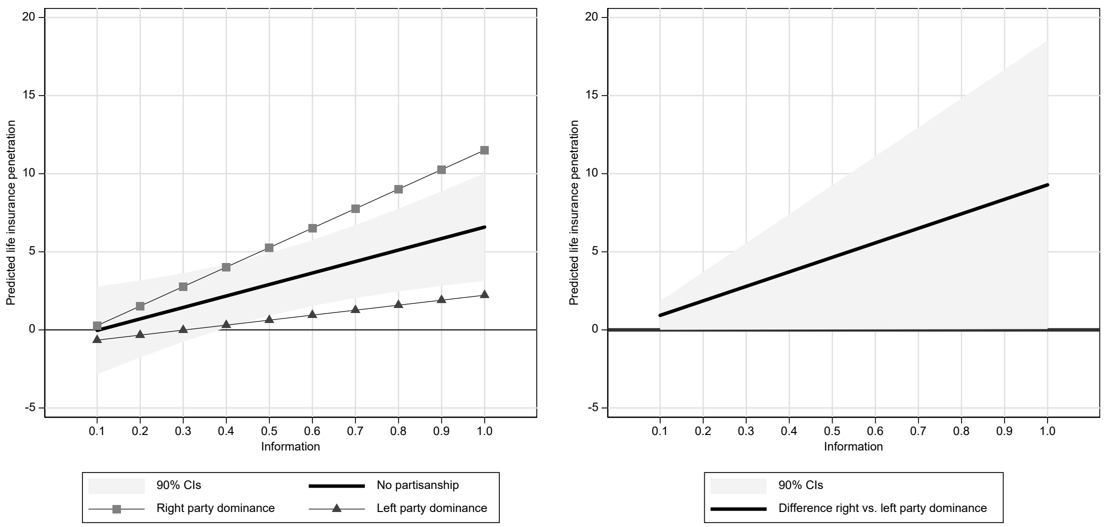
Our assembled dataset contains 486 country-year observations, covering 22 countries over the period from the early 1980s to the 2010s. Year coverage varies by country, giving us an unbalanced cross-section time series dataset. To emphasize the dynamic nature of our argument and data, we estimate an error correction model (ECM) with panel-corrected standard errors and with an AR1 autocorrelation structure. The estimation results are illustrated in Figure 1 (detailed results can be found in the Appendix). We find that going from the lowest to the highest level of information (0 to 1) raises life insurance market penetration by an average of about 4.5 percent in the first year, and by about 8 percent in the long run. This substantive effect is indicated by the solid upward sloping line in the left panel of Figure 1.

The left panel also includes separate estimates for countries with frequent left and right governments (H2b). The effect of information on life insurance penetration is much stronger in countries with frequent right governments (p95, top line), whereas it is muted in countries with frequent left governments (p5, bottom line). The right panel of the figure shows the difference between these lines (also based on Model (3)), which is substantive at high levels of information, and statistically significantly different from zero throughout. Left governments (rightly) worry about the expansion of life insurance markets largely because they could be a trojan horse for the expansion of private health insurance. While there are no

[11] The assumption is that if good diagnostics is a necessary condition for treatment; *ipso facto* effective treatment (fewer years of life lost) is a sufficient condition for accurate diagnostics. We realize that this will be a noisy indicator since some diagnoses may not be followed by treatment, and some treatment may make more effective use of information. But as long as the variance is not systematically related to our dependent variable (market development) it will only bias our results toward zero.

readily available data on the regulation of the life insurance industry, a typical restriction is that insurance companies cannot use genetic information to set premiums. This creates an adverse selection problem for insurers.

Figure 1: Predicted life insurance penetration



Note: Simulations are based on Models (1) and (3) in Table 2 in the Appendix.

The quantitative results are clearly only suggestive—we do not claim to have identified causal effects—but they lend credence to the proposition that the increased availability of diagnostic testing has facilitated life insurance markets, barring regulations designed to counter this trend.

Conclusion

Are the developments in the private insurance market—fueled by the information revolution—a harbinger of what is to come in social insurance? Will the information revolution undermine majority support for public social policy programs, lead to widespread privatization, and end the welfare state as we know it? Our argument implies that as more information can be credibly shared with insurers, private markets in social insurance are becoming more feasible. However, as mentioned, pressure to marketize is not the same as privatization, and some of the following factors are likely to prevent private insurance, even in the presence of extensive and symmetric information.

First, markets can be preempted by allowing greater differentiation in the public sector. We have not presented data to explore this hypothesis, but there is ample evidence that this is in fact the case, albeit to different degrees depending on government partisanship (Gingrich 2011; Hacker 2004). Second, pressures to marketize will be tempered by governments using measures such as "price non-discrimination" clauses to rule out private markets. These can prevent insurers from offering better deals to those with lower-risk profiles. Left governments are more likely to pursue such policies because they represent constituencies that are lower income and higher risk. We find support for this conjecture in the literature.

When markets are blocked for any reason, increased information will tend to polarize preferences regarding the level of public provision and the distribution of the costs. This polarization can intensify in the presence of effective private options because such options make the feasibility of markets clear—even if it is hard for individuals to take advantage of the options because of the double payment problem.

Third, correlated risk can undermine private insurance markets because insurers cannot rely on average risks to set premiums that equal payouts in expectation. Consequently, in areas where correlated risks are typical—such as the domain of unemployment[12]—social insurance may continue to be the only

[12] There are, however, unemployment systems that have features of private markets (Sweden is a good example).

feasible option. In these cases, increasing information should lead to more polarized attitudes (and demand for segmentation within a public system), but not to privatization.

Fourth, time-inconsistency problems arise in some insurance domains, and the state is able to address these more credibly than private companies. In commercial insurance, younger, healthier, and more employable workers "vote with their feet" by simply leaving the insurance scheme. This is analogous to the aforementioned adverse selection problem, but the difference is that these workers would prefer to stay and would do so if the insurer could make credible commitments to future benefits. Private insurers are not generally in a strong position to be able to solve this problem because they lack coercive powers, and instead they focus on product markets where the problem does not arise in the first place.

In homogenous risk pools, the time-inconsistency problem is muted for risks that are fairly stable over time, and where people continuously buy insurance so that they are covered for the next insurance term. For example, home insurance covers risks that usually do not change much from one contract period to the next, and consequently there is no implied transfer (in expectation) to others. Time-inconsistency becomes a serious problem when large transfers are required across generations because younger generations know that they are unlikely to need the insurance until they grow old. This is obviously true of PAYG old-age insurance, but it is also true of health insurance because poor health is more prevalent among the old. If insurance companies offered health plans that insured people indefinitely, they may well maximize the lifetime utility of would-be buyers, but young, healthy people would not buy into the insurance unless they could be certain that it would pay out when they grew old and sick, and unless insurers could assess long-term risks accurately there would, again, be a standard adverse selection problem. The time-inconsistency problem suggests that some types of insurance may continue to only be provided through the state.

Fifth, switching from social to private insurance may incur considerable costs, especially for those that have contributed to the public system for an extended period of time. While these individuals may be better off in a private system, they may still continue to support the public system because they are already vested in it. For this reason alone, the transition from social to commercial insurance is likely to be gradual, even in the presence of breathtaking technological advances regarding information collection and sharing.

Finally, there is no private insurance against poverty, such as in the case of chronic illness. Many middle-class Americans deplete their savings after exhausting private insurance to pay for long-term care because of Alzheimer's and other debilitating diseases. In such instances, the individuals concerned will eventually qualify for Medicaid. Even though it targets the poor, Medicaid enjoys broad support among the middle class who tends to be privately insured (Busemeyer/Iversen 2020). A similar argument applies to the state as a backstop for insurance bankruptcies.

Market failure is the starting point for almost all research on social insurance—which is often treated as synonymous with the welfare state—but if the source of market failure is incomplete information then surely more information will change the politics of social insurance. Based on our theoretical framework, the information revolution will lead to increased polarization of welfare state politics. Moreover, at least in some areas of social insurance, it might also lead to the (gradual) privatization of currently socialized risk pools. Either way, solidaristic risk-sharing will come under political attack, and inequality in social protection and incomes is likely to increase.

Appendix

Life insurance penetration data:

The data are available at: https://stats.oecd.org/Index.aspx?DataSetCode=INSIND (OECD insurance indicators). See also OECD (2015).

Sample: AUS (1984–2013), AUT (1987–2013), BEL (1983–2013), CAN (1984–2013), CHE (1983–2013), DEU (1987–2013), DNK (1983–2013), ESP (1983–2013), FIN (1983–2013), FRA (1983–2013), GBR (1996–2013), GRC (1992–2013), IRL (1983–2013), ISL (1983–2013), ITA (1983–2013), JPN (1983–2013), NLD (1995–2013), NOR (1983–2013), NZL (1989–2003), PRT (1983–2013), SWE (1983–2013), USA (1983–2013).

Mortality by cause data:

Our measure of information is based on data about premature mortality, as provided by the OECD (https://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT).

We make use of the "Potential Years of Life Lost" (PYLL) variable, which is defined as follows: "This indicator is a summary measure of premature mortality, providing an explicit way of weighting deaths occurring at younger ages, which may be preventable. The calculation of Potential Years of Life Lost (PYLL) involves summing up deaths occurring at each age and multiplying this with the number of remaining years to live up to a selected age limit (age 70 is used in OECD Health Statistics). In order to assure cross-country and trend comparison, the PYLL are standardized, for each country and each year. The total OECD population in 2010 is taken as the reference population for age standardization. This indicator is presented as a total and per gender. It is measured in years lost per 100 000 inhabitants (men and women) aged 0-69."

[Source: OECD (2015), Potential Years of Life Lost (indicator). doi: 10.1787/193a2829-en (retrieved September 1, 2015)].

We calculate Potential Years of Life Lost (PYLL) due to the following diseases:

- Certain infectious and parasitic diseases
- Neoplasms
- Diseases of the blood and blood-forming organs
- Endocrine, nutritional, and metabolic diseases
- Mental and behavioral disorders

- Diseases of the nervous system
- Diseases of the circulatory system
- Diseases of the respiratory system
- Diseases of the digestive system
- Diseases of the skin and subcutaneous tissue
- Diseases of the musculoskeletal system and connective tissue
- Diseases of the genitourinary system
- Certain conditions originating in the perinatal period
- Congenital malformations and chromosomal abnormalities

These diseases account for about 75 percent of PYLL—the remaining PYLL are largely due to “external causes of mortality” (traffic accidents, accidental poisoning, suicides, etc.).

The WHO data rely on the International Statistical Classification of Diseases and Related Health Problems (ICD). Over time, the ICD has been updated. In the empirical analyses, we include an indicator variable for changes to the ICD classification. The potential breaks occur in the following country-years: Australia (AUS): 1968, 1979, 1998. Austria (AUT): 1969, 1980, 2002. Belgium (BEL): 1968, 1979, 1998. Canada (CAN): 1969, 1979, 2000. Denmark (DNK): 1969, 1994. Finland (FIN): 1969, 1987, 1996. France (FRA): 1968, 1979, 2000. Germany (DEU): 1998. Greece (GRC): 1968, 1979. Iceland (ISL): 1971, 1981, 1996. Ireland (IRL): 1968, 1979, 2007. Italy (ITA): 1968, 1979, 2003. Japan (JPN): 1968, 1979, 1995. Luxembourg (LUX): 1971, 1979, 1998. Netherlands (NLD): 1969, 1979, 1996. New Zealand (NZL): 1968, 1979, 2000. Norway (NOR): 1969, 1986, 1996. Portugal (PRT): 1971, 1980, 2002. Spain (ESP): 1968, 1980, 1999. Sweden (SWE): 1969, 1987, 1997. Switzerland (CHE): 1969, 1995. United Kingdom (GBR): 1968, 1979, 2001. United States (USA): 1968, 1979, 1999.

Partisanship variable and controls:

Since partisanship only has an effect through slowly changing regulatory measures, we use Huber and Stephens' (2001) cumulative measure of seats in government held by left parties (divided by the number of years), starting in 1960. This measure does not change a great deal in our sample, and since we estimate fixed effects models, it is not clear that estimating direct effects of partisanship is meaningful. We report results for including left partisanship both as an independent variable and as an interaction with information, but our focus is on whether left partisanship slows the progression of markets in response to information, as hypothesized above.

Finally, we include three control variables that may influence life insurance penetration: (i) the percentage of the population covered by public or primary private health insurance; (ii) total health expenditure (all financing agents) as a % of GDP; and (iii) the rate of economic growth.

Sample:

The following country-years are included in our sample, which was determined by data availability: AUS (1985–2011), AUT (1988–2013), BEL (1984–2012), CAN (1985–2011), CHE (1984–2012), DEU (1993–2013), DNK (1984–2012), ESP (2011–2011), FIN (1984–2013), FRA (1991–2011), GBR (1997–2013), GRC (1993–2008), IRL (1984–2010), ISL (1984–2009), ITA (1989–2012), JPN (1984–2012), NLD (1996–2013), NOR (1984–2013), NZL (1990–2001), PRT (1984–2013), SWE (1986–2013), USA (1984–2010).

Results:

The estimation results are shown in Table 2. The only difference between the models is whether and how cumulative partisanship is entered as an explanatory variable. Model (1) does not contain a partisanship variable, while Model (2) does—both as an independent variable and as an interaction with information. Finally, Model (3) only includes partisanship as an interaction term. Because the cumulative partisanship variable does not change very much within countries over time, it provides little information that is not already contained in the fixed effects. The interaction, on the other hand, tests whether the effects of changes in information are conditioned by stable differences in partisanship, the key question for our partisan hypothesis. The controls are health insurance coverage (% of population); total health expenditures (% of GDP); economic growth rate; partisanship; country dummies; and an indicator variable for potential breaks in the PYLL series. We adjust the variance-covariance matrix in the reported results by applying the correct mean squared error (Baltagi 2011).

Table 2: Life insurance penetration, information, and partisanship (ECM)

	(1)	(2)	(3)
	Dependent variable: Life insurance penetration [first difference]		
Life insurance penetration [lag]	-0.185** (0.066)	-0.218** (0.068)	-0.199** (0.067)
Information [lag]	1.357* (0.597)	2.968** (1.064)	2.560* (1.058)
Information [first difference]	2.789+ (1.578)	7.729+ (4.426)	6.322 (4.830)
Left partisanship X Information [lag]		-0.036** (0.014)	-0.029* (0.015)
Left partisanship X Information [first difference]		-0.151 (0.106)	-0.100 (0.111)
Left partisanship		0.031** (0.009)	
Health insurance coverage (% of population)	-0.012 (0.023)	-0.018 (0.023)	-0.010 (0.023)
Total health expenditures (% of GDP)	-0.059 (0.066)	-0.034 (0.060)	-0.059 (0.066)
Economic growth rate	0.036+ (0.020)	0.035+ (0.020)	0.038+ (0.021)
Constant	2.011 (2.299)	1.208 (2.166)	1.951 (2.304)
Dummy for breaks in PYLL series	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes
N	500	500	500
N of countries	22	22	22
Adj. R2	0.127	0.148	0.137

Note: Coefficients above SEs. + p<0.10, * p<0.05, ** p<0.01

The results for information are very similar across all three models: the lagged dependent variable is statistically significant, with an estimated coefficient of around -0.2, and the lagged information variable is positive and statistically significant (indicating that more information is associated with higher life insurance penetration). The coefficients on the regressors have very specific interpretations: the coefficients on the lagged level variables capture permanent effects of a one-off change in those variables, while the coefficients on change variables capture transitory effects (Beck/Katz 1995). We find that there are short-term transitory effects (not statistically significant), but that the main effects are long term and permanent.

References

Akerlof, George A. 1970. “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism.” *The Quarterly Journal of Economics* 84(3):488–500.

Baldwin, Peter. 1990. *The Politics of Social Solidarity: Class Bases of the European Welfare State, 1875-1975*. Cambridge: Cambridge University Press.

Baltagi, Badi H. 2011. *Econometrics*. Berlin: Springer Science & Business Media.

Barr, Nicholas. 2012. *Economics of the Welfare State*. Oxford: Oxford University Press.

Beck, Nathaniel, and Jonathan N. Katz. 1995. “What to Do (and Not to Do) with Time-Series Cross-Section Data.” *American Political Science Review* 89(3):634–47.

Boadway, Robin, and Michael Keen. 2000. “Redistribution.” Pp. 679–789 in *Handbook of Income Distribution*, edited by A. Atkinson and F. Bourguignon. Amsterdam: Elsevier.

Busemeyer, Marius R., and Torben Iversen. 2020. “The Welfare State with Private Alternatives: The Transformation of Popular Support for Social Insurance.” *The Journal of Politics* 82(2):671–86.

Coletti, Margaret H., and Howard L. Bleich. 2001. “Medical Subject Headings Used to Search the Biomedical Literature.” *Journal of the American Medical Informatics Association: JAMIA* 8(4):317–23.

Esping-Andersen, Gøsta. 1985. *Politics Against Markets: The Social Democratic Road to Power*. Princeton: Princeton University Press.

Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Princeton: Princeton University Press.

Gingrich, Jane. 2011. *Making Markets in the Welfare State: The Politics of Varying Market Reforms*. Cambridge: Cambridge University Press.

Hacker, Jacob S. 2004. “Privatizing Risk without Privatizing the Welfare State: The Hidden Politics of Social Policy Retrenchment in the United States.” *American Political Science Review* 98(2):243–60.

Huber, Evelyne, and John D. Stephens. 2001. “Welfare State and Production Regimes in the Era of Retrenchment.” Pp. 107–45 in *The New Politics of the Welfare State*, edited by P. Pierson. Oxford: Oxford University Press.

Iversen, Torben, and David Soskice. 2001. “An Asset Theory of Social Policy Preferences.” *American Political Science Review* 95(4):875–95.

Korpi, Walter. 1983. *The Democratic Class Struggle*. London: Routledge.

Moene, Karl O., and Michael Wallerstein. 2001. “Inequality, Social Insurance, and Redistribution.” *American Political Science Review* 95(4):859–74.

National Institute for Health Care Management. 2012. “The Concentration of Health Care Spending.” NIHCM Foundation Data Brief (July).

OECD. 2015. *OECD Insurance Statistics 2014*. Paris: OECD Publishing.

Pagana, Kathleen D., and Timothy J. Pagana. 1992. *Mosby’s Diagnostic and Laboratory Test Reference*. Amsterdam: Elsevier.

Pagana, Kathleen D., Timothy J. Pagana, and Theresa N. Pagana. 2014. *Mosby's Diagnostic and Laboratory Test Reference*. Amsterdam: Elsevier.

Przeworski, Adam. 2003. *States and Markets: A Primer in Political Economy*. Cambridge: Cambridge University Press.

Rehm, Philipp. 2016. *Risk Inequality and Welfare States. Social Policy Preferences, Development, and Dynamics*. Cambridge: Cambridge University Press.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics* 90(4):629–49.

Shojania, Kaveh G., Elizabeth C. Burton, Kathryn M. McDonald, and Lee Goldman. 2003. "Changes in Rates of Autopsy-Detected Diagnostic Errors over Time: A Systematic Review." *JAMA* 289(21):2849–56.

Siegel Bernard, Tara. 2015. "Giving Out Private Data for Discount in Insurance". *The New York Times*, April 8.

Singer, Natasha. 2017. "How Big Tech Is Going After Your Health Care." *The New York Times*, December 26.

Stephens, John D. 1979. *The Transition from Capitalism to Socialism*. London: Macmillan.

Stiglitz, Joseph E. 1982. "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics* 17(2):213–40.

de Swaan, Abram. 1988. *In Care of the State: Health Care, Education and Welfare in Europe and the USA in the Modern Era*. Oxford: Oxford University Press.