The market for creampuffs: Big Data and the transformation of the welfare state

Torben Iversen

Harvard University

Philipp Rehm

Ohio State University

(First draft, June 2015. This draft: April 2016)

Abstract:

The literature on the welfare state assumes, often implicitly but almost universally, that social insurance can or will be provided through the state. This assumption is based on economic models of insurance that show the propensity for market failure when information is limited and privately held. With the data revolution this is no longer a satisfactory approach, and this paper asks what happens when information rises and can be credibly shared with insures. Our model shows that Big Data alters the politics of social insurance by increasing polarization over the level and cost-sharing of public provision, and sometimes by creating majorities for a shift towards segmented and inegalitarian private markets (a shift that is conditioned by government partisanship). We offer a preliminary test of the model examining the relationship between information and life insurance market penetration and between information and polarization.

Introduction

A central function of the welfare state is to provide social insurance. Indeed, a voluminous literature underscores that political support for social spending in the pivotal middle and upper-middle classes rests on their demand for insurance against risks like unemployment, illness, and old age (e.g., Esping-Andersen 1990; Iversen and Soskice 2001; Moene and Wallerstein 2001; Rehm 2011b). Underlying most of these analyses is an assumption, often implicit but virtually universal, that social insurance cannot be provided effectively through private markets due to incomplete and asymmetric information (see Barr 2012, Boadway and Keen 2000 for standard treatments).[1] While this assumption may have been approximated in the past, the data revolution is making it untenable. This paper asks what happens to the welfare state when information about health, unemployment risks, life expectancy, credit-worthiness, and so on, becomes more widely available and shareable? Theory and evidence suggest that this will have consequences for the politics of social protection, and we may in fact be in the early phases of a major transformation.

A hint of what is to come can be gleaned from the car insurance market. There are now more than a dozen insurance companies in the US offering "pay as/how you drive" policies which tie premiums to driving behavior (based on a black box that tracks distance, acceleration, braking, cornering forces, speeds relative to speed limits, among other things). Unsurprisingly, safe drivers flock to these plans. There are also anecdotes hinting at the future of life and health insurance. For example, John Hancock Life Insurance, a major player in the American market, introduced a policy that calculates annual premiums partially based on data collected by an "activity tracker" that policy-holders receive for free when they sign up. These types of devices (which include smartphones equipped with the right app) can track and instantly share things like: steps and stairs taken, active minutes, calories burned, heart rate, sleep quality, blood pressure, among others. The company markets 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."[2]

[1] In a rare acknowledgement of this omission, Moene and Wallerstein (2001, 871) note in a widely cited paper that "[t]heoretically, the largest gap in our approach is the absence of a private alternative to publicly provided insurance. … The politics of the demand for insurance when there is a private alternative involves different considerations."

[2] FAQs on the John Hancock Vitality Program (http://jh1.jhlifeinsurance.com/jhl-ext-templating/filedetail?vgnextoid=96b80f7ed8f0d410VgnVCM1000003e86fa0aRCRD&siteName=JHSalesNet). The 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" (http://www.nytimes.com/2015/04/08/your-money/giving-out-private-data-for-discount-in-insurance.html?_r=0).

To understand the consequences of "Big Data" (massive data that can be credibly shared and quickly analyzed), we take as our starting point Akerlof's classic model of "the market for lemons" (Akerlof 1970), with notable extensions by Rothschild and Stiglitz (1976), and Stiglitz (1982). The standard model provides a powerful logic to justify public provision in areas such as health insurance, but it completely ignores the politics of public provision. The standard model also shows that information can make private insurance markets a viable alternative to state provision, but information plays no role in the contemporary comparative political economy of social protection. We therefore extend the standard insurance model to show how information about social risks affects the politics of social protection.

Our main thesis can be summarized in two key propositions: First, the data revolution expands the range of social protection that can be provided privately, which undermines cross-class support for the welfare state and may result in an expansion of private markets. Left governments that represent more risk-exposed individuals may slow this change, but the growing feasibility of market solutions puts pressure on politicians of all stripes to permit markets to work. Second, even when social protection is provided through the state in areas where private provision is not feasible, the structure of political support will become more class-polarized as information spreads.

Our primary aim is analytical, but we offer two preliminary tests of our argument. First, we consider the expansion of private markets in life insurance. This is an area that works very similarly to health insurance, but where public provision is historically weak. We show that the growing availability of health information facilitates the expansion of such markets, but that this relationship is weaker in countries with a historically strong Left. For the second test we explore a policy area of insurance that, for reasons unrelated to information, has everywhere remained almost entirely public: unemployment insurance. We show that improved individual information about risk is associated with a polarization in the support for public insurance, even after controlling for the variance in actual risk. As the veil of ignorance is lifted, voters become more divided and solidaristic solutions become harder to engineer.

2. State of the literature

Akerlof's seminal 1970 *QJE* article lays out a key reason for the breakdown of markets: private information and the associated problem of adverse selection. Rothschild and Stiglitz (1976) and Stiglitz (1982) develop the logic for insurance markets, with more recent extensions summarized in Boadway and Keen (2000), Barr (2012, chap. 4), and Przeworski (2003, chap. 11). In these models individuals know their risk-types, but insurers do not. Market failure occurs when plans that are directed at those with low risks are picked up by those with high risks, raising prices, and driving out the good risks. In Akerlof's model, "adverse selection" results in a market only for bad risks, or "lemons."

Less stark conclusions follow when people are risk-averse and therefore willing to buy relatively "expensive" plans that include some bad risks, and when insurers can rationally anticipate that targeting low-risk types ("cream skimming") cannot be done without attracting high-risk types. A "pooling" equilibrium is one in which relatively low- and high-risk types are offered a single insurance plan at a single price. We show in the next section that such equilibria can exist, but that they are inefficient because some low-risk types will not buy insurance, which raises the price for others and makes insurance unaffordable for some.[3]

In addition to asymmetric information, private insurance markets may be underdeveloped because of correlated risks. Insurers can act as risk-neutral profit maximizers only when risks in the insurance pool cancel each other out, implying that the number of insurance claimants is fairly constant. If risks are highly correlated, insurers can face insurmountably large insurance claims in the case of a major shock. This is one obvious reason that unemployment insurance is almost invariably state-provided or heavily state-subsidized and regulated.

There are three limitations to the standard economic analysis. First, and most obvious, it does not capture the *politics* of social insurance. While economists are usually contend to propose that market failure functionally requires (and justifies) government intervention, public provision will be opposed by those with the lowest risks – such as the young and healthy in the case of healthcare. They may prefer to go entirely uninsured in a private market to paying into a public system where they heavily subsidize people at high risk. Yet, for the same reason a political majority may prefer such a system as way of reducing their own costs. When they do, majority demands become the driver of public provision.

Second, the assumption in standard models that risk is only known to buyers is increasingly untenable. Low-risk types have an incentive to share their information with insurers to reduce their premiums, and better access to independent collection and validation of information is facilitating such sharing. Once credible information sharing is feasible, "separating equilibria" with highly segmented risk markets can exist. Moreover, if markets are feasible a majority may prefer a private system. This is true if we assume that private insurance is not less cost-effective

[3] Appendix A considers the possibility of a "separating equilibrium" with asymmetric information because insurers can monitor individual purchases of insurance.

than public provision, and that the risk-distribution is right-skewed. The latter is normally the case (see Rehm 2016).

Third, the standard model focuses on asymmetric information without considering the effects of differences in the level of information. Even when risks such as unemployment are uninsurable in private markets, information will shape the structure of support for public spending. Since those with low risks subsidize those with high risks, public risk pooling is always strongly redistributive. But it is only when the risks are known that this becomes a source of political division. More information therefore also means more conflict and less cross-class "solidarity". Again, information has political consequences.

Our focus on information sets our argument apart from most existing literature on the politics of social insurance. This work has focused on how peoples' position in the economy affects their exposure to risk and consequently their preferences for insurance. Cameron (1978), Garrett (1998), Rodrik (1998), and Wren and Rehm (2014) show how economic openness can be a source of risk, and hence preferences; Esping-Andersen (1990, chap. 8) and Iversen and Cusack (2000) consider the consequences of deindustrialization for risk and preferences; Mares (2003) examine such differences across industrial sectors; Moene and Wallerstein (2001) look at the relationship between income and demand for insurance; Iversen and Soskice (2001) introduce the role of skills for risk-exposure and Rueda (2007) considers divisions between secure insiders and insecure outsiders; Rehm (2011b, 2016) and Alt and Iversen (forthcoming) consider more broadly how the distribution of risk affects the structure of social policy preferences. Strikingly, none of this work considers information as an independent causal variable explaining both preferences and size of the public system. The focus is exclusively on determining who is at risk and how this shapes interests.

Because political economy models of social insurance have ignored the issue of information, we will spend some time introducing the classic insurance model and show how it can be extended to the political world that is the focus of work in political science. We will show that information transforms the political economy of insurance in ways that are not accounted for in the existing literature.

3. The model

We present the argument in three steps. We begin by introducing the classic asymmetric information case and show that it typically leads to majority support for public provision, although (inefficient) markets for insurance are feasible. We then turn to the symmetric information case and show that when information is plentiful and can be credibly shared with insurers, a majority may prefer private provision. When market solutions are blocked, whether for political or economic reasons, preferences over public provision will polarize. Finally, we consider the role of government partisanship.

3.1. Basic setup

We assume that people are looking one period into the future and decide how much of their current income to spend on insurance against risks of losing that income in the same period (as a result of illness, unemployment, etc.).[4] The model uses log utility to capture risk-aversion (RRA=1).[5] Specifically, the (von Neumann–Morgenstern) expected utility of individual *i* is defined as:

(1) $$U_i = \ln\left(y_i - c_i\right)\cdot(1-p_i) + \ln\left(k_i + b_i\right)\cdot p_i ,$$

where $y_i > 1$ is income when in the good state, c_i, is the cost of insurance, p_i, is the risk of losing income, b_i is the benefit in the bad state, and k_i is a private, pre-transfer income when in the bad state. If the bad state is one in which the individual is unable to work, k_i can be understood as non-labor income from savings or other assets ("self-insurance").

We initially assume that *i* knows everything there is to know: y_i, c_i, p_i, k_i, and b_i. If the insurer also had this information it could offer insurance plans for each risk group, using premia by group members in the good state to pay for benefits of those in the bad state. If c_i is defined as a share, π_i, of income that goes to paying for insurance, so that $c_i = \pi_i \cdot y_i$, and if we ignore administrative costs of provision as well as any markups, which are irrelevant for our results, the insured individual benefit is:[6]

(2) $$b_i = \frac{\pi_i \cdot (1-p_i)}{p_i}\cdot y_i ,$$

where $\pi_i \cdot (1-p_i)/p_i$ is the income replacement rate. The micro logic is illustrated in Appendix A in an example with three risk groups. The ratio $-(1-p_i)/p_i$ is the slope of the "fair-bet" line, which is the line that makes expected income in the two states identical. This line is also the set

---

[4] The length of the period is not important to or argument. For health insurance, for example, it may be quite long.
[5] We ignore more complicated utility functions in the interest of presentational economy. The key results only require that RRA>0.
[6] To see this, note that the total amount paid out by the insurer to a group of N members with risk p_i is $N \cdot p_i \cdot b_i$ while the total premium received is $N \cdot (1-p_i)\cdot \pi_i \cdot y_i$. The insurer breaks even when the two equals each other, which gives (2).

of combinations where the insurer breaks even, assuming that risks are pooled across a large group of individuals with uncorrelated risks.

Since i is risk-averse, he or she will purchase enough insurance to equalize expected income across the two states, which is the value of π_i that maximizes (1):

$$\pi_i^* = p_i \cdot \left(1 - \frac{k_i}{y_i}\right) = p_i \cdot (1 - s_i), \tag{3}$$

where s_i is non-labor income in the bad state as a share of labor income in the good state. Unsurprisingly, the higher the risk of losing income the greater the share of income spent on insurance, and the more non-labor income in the bad state the lower the demand for insurance in the good state. Higher labor income increases demand for insurance, but higher non-labor income reduces demand. If the latter comes from savings out of the former, demand will be declining in the savings rate, which tends to be rising in income. Since this is not important to our results we simply assume that the savings ratio is constant: $s_i = s$.[7]

3.2. The asymmetric information case

Insurance markers are not problematic if individual risks are observed by insurers. As long as there are large groups with the same risk, and individual risks are not correlated, each individual would pay π_i^* in the good state and get a corresponding benefit of b_i in the bad state. But if insurers do not have individual information about risks, they will only know the mean risk in the pool of insured, p, which can be observed as the share of insured individuals who collects insurance.

With asymmetric information the benefit received by i now depends on p instead of p_i:

$$b_i' = \frac{\pi_i \cdot y_i \cdot (1-p)}{p}, \tag{4}$$

and the preferred level of insurance is a function of both individual risk and average risk: [8]

---

[7] Recent work by Ansell (2014) and others emphasize the role of non-labor income for preferences, and our model is entirely consistent with this work. But non-labor income is not the focus of this paper.

[8] This is the value of π_i that maximizes equation (1) when the replacement rate is $\pi_i \cdot (1-p)/p$ instead of $\pi_i \cdot (1-p_i)/p_i$.

(5) $\pi_i^{**} = p_i - (1-p_i)\cdot s \cdot \frac{p}{1-p}$.[9]

At this price and replacement rate, will individuals buy insurance? The answer depends on their own risk relative to the risk of others. Those with p_i above *p* will find the pooled insurance plan an unambiguously good deal; those with a p_i below *p* may or may not. This is because for these people there is an additional cost of social insurance, which is the implied subsidy to those at higher risk. This cost is a function of the composition of risk in the pool of insured, so each individual's decision to buy insurance depends on the decision of others. We can model this as the outcome of a Schelling network game (Schelling 1978).

Specially, an individual will buy (some) insurance if:

$$\pi_i^{**} > 0$$

(6) $\Downarrow$

$$p_i > \frac{s}{1/p - 1 + s}.$$

For example, if everyone is insured and the mean risk is .25 and *s*=.5 (so that people can sustain half of their income in the bad state), those with risks below 14 percent would be below the threshold and give up their insurance. Then the average risk in the insurance pool would rise and more people would exit, and so on. This is Akerlof's lemons logic at work; the process is shown in a step-by-step graphic in Appendix A.

Yet, this logic does not end in a market for lemons only. The reason is that there are always some (except for a limiting case) below the highest risk who are themselves sufficiently exposed that they will be willing to pay for an insurance that includes higher risks. The equilibrium is found where the expected mean risk in the insured population is equal to the actual mean risk:

---

[9] Note that equations (4) and (5) imply that individuals can buy as much insurance as they like, and that their benefit is commensurable with that payment. Specifically, insurance companies take in an average of $\pi \cdot y$ per insurance holder, and they pay out an average of $\pi \cdot y \cdot \frac{(1-p)}{p}$ for each insurance claimant, but each individual is paid in proportion to their insurance payment: $\frac{\pi_i \cdot y_i}{\pi \cdot y} \cdot \pi \cdot y \cdot \frac{(1-p)}{p} = \pi_i \cdot y_i \cdot \frac{(1-p)}{p}$ (equivalent to (4)). So each insurer knows the amount of insurance bought from them by each individual. But they do not know the total amount of insurance bought by each individual from all insures, nor do they know whether the premium paid to them is due to the level of individual risk, to individual income, or to the individual degree of risk-aversion. This prevents any insurer from offering a "cheap" plan that will only be taken up by low-risk types. In our model this makes a pooling equilibrium possible. The logic is illustrated in detail in Appendix A.

$$p^{p_i > c} = p^e,$$

where $p^{p_i > c}$ is the mean risk in a pool of people who are above a critical threshold, *c*, which is given by equation (6) above (where $p = p^e$). The equilibrium will depend on the distribution of critical thresholds which is determined by *s* (the more self-insurance, the lower the threshold) and the distribution of risk, p_i (roughly, the "thinner" the right tail, the more will buy insurance).[10]

The logic is illustrated in Figure 1, which uses an example where *s*=.5 and the distribution of risk is even in the interval [0, .5]. If everyone buys into the insurance plan, so that coverage is 100 percent (recorded on the right *y*-axis), the expected *p* is .25. At this implied "price" only those with individual risks above .14 (given by equation 6) would buy insurance, which is equivalent to 71 percent of the population. In this case, the observed average risk is .33 (the average for those with critical thresholds above .14). This logic continues until the curve that maps expected onto actual risks intersects the 45-degree line. At this point the equilibrium of the game is reached, which in our example implies that a small majority (57 percent) of the population buys private insurance.[11] So Akerlof's conclusion was too gloomy.[12]

Whatever the outcome in the game, since *someone* at low risk will always find it preferable to opt out of a private insurance plan in a world of imperfect information – even though they would buy insurance in a world with complete information – it is inefficient. This is where the economic analysis ends, and public provision is "explained" as a solution to this problem of under-provision. But efficiency in not what brings a public system about politically. If the median voter sets the tax rate in such a system, which is equivalent to choosing π for everyone, the low-risk minority will subsidize the majority. Since we know these people would forgo private insurance when they can choose their own contribution level, it follows that they would also not want a public system where they are forced to pay at the same rate as everyone else. Demand for a public system *does not* come from low-risk, uninsured people.

[10] In a more elaborate model it will also depend on the degree of risk-aversion.

[11] Note that depending on the distribution of risk there may be more than one equilibrium. This is not important for our purposes.

[12] Akerlof's market for only lemons emerges when the average-risk line meets the 45-degree line at the highest level of risk. This is always the case in Akerlof's model because he assumes that people are risk-neutral. Risk-neutrality is equivalent to a situation where people have exactly the same income in the good and bad states, so that *s*=1. From (6) the condition is now $p_i > p$, which means that only those with above-average risk buys insurance, so no one buys, except for lemons who are indifferent ($p_i = p$). A lemons-plus market thus requires that people are risk-averse.

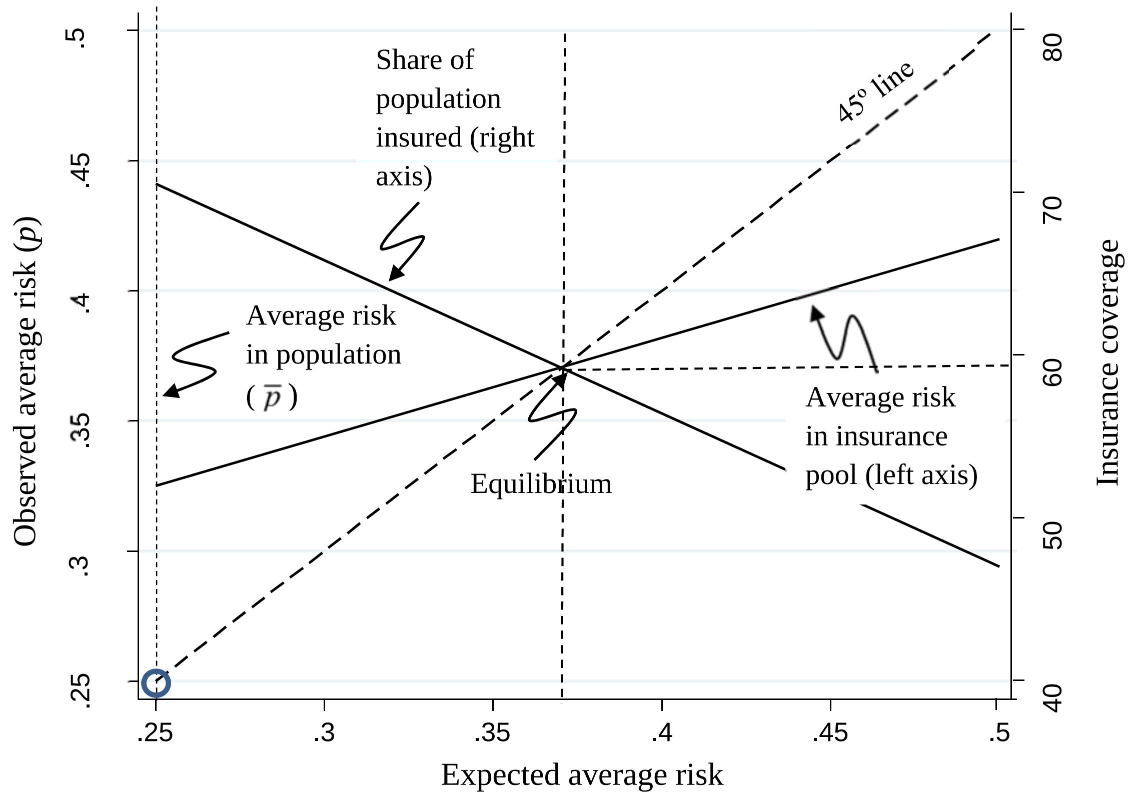
Instead, it comes from the majority who will be better off in a public system precisely because low-risk types now contribute to the pool. In other words, it happens for reasons of distributive politics, not efficiency. This is easy to see in our example where the median voter starts out being (partially) insured in the private market. In the specific example, the individual with the median risk (equal to .25) would spend 3.9 percent of her income on insurance in the private market and get a replacement rate of 6.7 percent in the case of income loss.[13] By contrast, in a public system with a proportional tax, the median voter would choose a 12.5 percent tax rate and get a 37.5 percent replacement rate, which would exactly equalize net income in the two states.[14] Equalizing income in the two states is optimal for risk-averse individuals (as we established above), and the private system clearly does not come close to this ideal. Information asymmetries thus explain why public insurance systems often enjoy widespread electoral support. At the same time is it clear that there will be conflict within the majority since those with lower risk prefer lower rates of taxation, and vice versa – an implication consistent with existing evidence (Cusack, Iversen, and Rehm 2006; Rehm 2009).[15]

[13] The median risk is .25 and the average risk in the pool is .36, when in equilibrium (see Figure 1). This gives the optimal spending from equation (5). The replacement rate can then be calculated from equation (4).

[14] Net income in the good state is $(1-.125)\cdot y_i = 875\cdot y_i$ and net income in the bad state is $(.5+.325)\cdot y_i = .875\cdot y_i$, (*s*=.325 and $\pi_i = .325$).

[15] It is also consistent with Moene and Wallerstein's (2001) conjecture that a means-preserving increase in inequality will reduce the median's preferred level of spending since insurance is a normal good in our model.

Figure 1: Equilibrium in the private insurance network game with private information



Note: Simulations assuming s=.5 and $p_i = [0, .5]$

3.3. The symmetric information case

Akerlof and Stiglitz did not discuss the case of symmetric information since they were interested in exploring the consequences of private information. But the symmetric information case is important to our story, and it comes in two varieties. The first is where *neither* buyers nor sellers have individual information about risk (low information). The second is where *both* do (high information).[16]

3.3.1. *Low information*. In the low information case, individuals will have to rely on the same aggregate information as insurers. Each person will have to form an expectation of their risk

[16] The set of case of asymmetric information we have covered is where only the individual has information; the set where only the insurer has information is presumed empty.

based on the observed number of unemployed, disabled, sick, and so on. In our model this is simply the mean risk in the population: $p_i^o = \bar{p}$, where p_i^o is i's observed level of risk and $\bar{p}$ is the overall population mean (as distinct from the mean among the insured, p). Since everyone has the same expectation, including the insurer, and since $\bar{p} \cdot (1-s)$ is everyone's preferred spending (from equation 3) $p = \bar{p}$ is an equilibrium (i.e., the risk in the insurance pool equals the mean risk in the population). In Figure 1 this special case is marked with a circle at the bottom left corner (which, it will be noted, is on the 45-degree line and therefore sustainable).

In the real world there may be no examples of such a complete lack of information, but the case is instructive nonetheless. The reason is that uncertainty reduces the variance in policy preferences: whereas the range of preferences in the private information case is $p_i = [p_{\min}, p_{\max}]$, in the case of uncertainty the range is $p_i^o = [p_{\min}^o, p_{\max}^o]$ (where, again, p_i^o is observed risk). The latter will be narrower than the former, which can be understood in a simple Bayesian framework. If *i* receives signals about his or her true risk, p_i, from a noisy environment in which the overall mean is $\bar{p}$ then:

(6) $$p_i^o = \alpha \cdot p_i^s + (1-\alpha) \cdot \bar{p}$$

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 the signals, which in our model equals the private information available to *i*.[17] As noted above, in the extreme case of no information ($\alpha = 0$) $p_i^o = \bar{p}$, so the range is zero; at the extreme of complete information $p_i^o = p_i$, the range equals the difference between those with the lowest and highest risk. A very simple way of stating the general insight is that *class conflict increases with information*. This echoes Rehm's (2011b) insight 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.

3.3.2. *High information*. The final, and important, case is where information is plentiful and can be shared between buyer and provider. Privacy legislation may limit the ability of insurers to acquire individual information, especially in the case of medical records, but insurers may not have to unlock private information – people may choose to give it to them voluntarily.[18] The examples in the introduction of using monitoring devices to reduce insurance premia suggest why.

[17] A formal proof can be found in Iversen and Soskice (2015), Appendix B

[18] We do not mean to minimize the capacity of firms to obtain data without peoples' consent. Sweeney (2002) shows the possibility in data rich environments of using anonymous information to identify individuals and thus in effect make private information public. But such unsolicited data acquisition is not necessary for our argument.

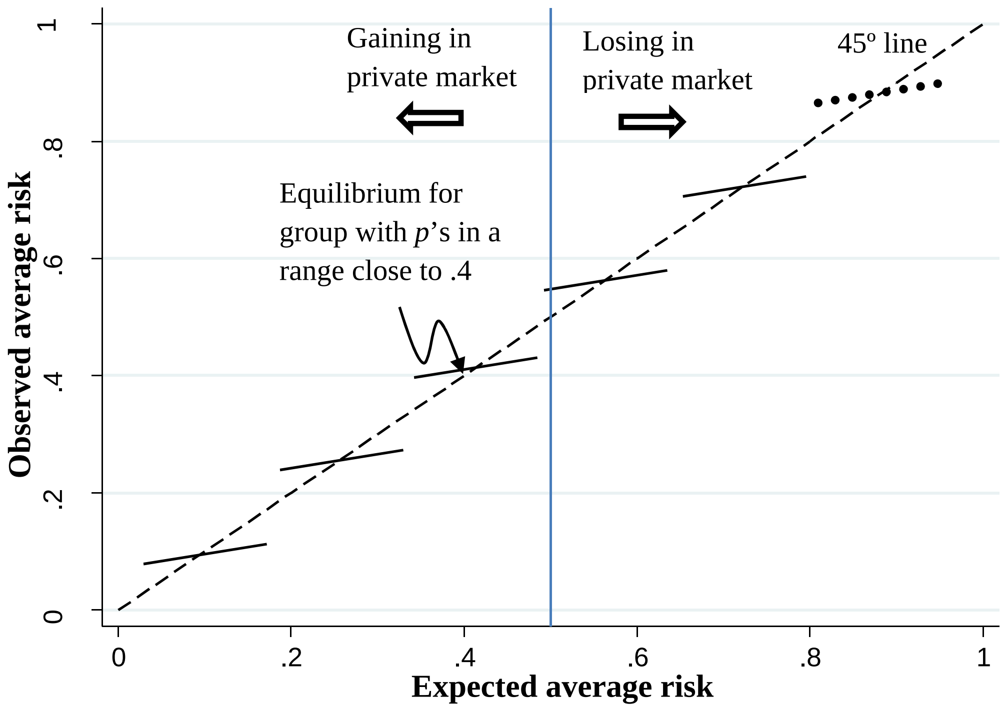
This logic applies to the important area of private health data where the level and credibility of information has vastly improved. There are two related forces behind this trend. First, the general advance of medicine has made diagnostics much more reliable (Shojania et al. 2003). Second, the explosion in the number and variety of tests that can be done by certified labs has made it possible to share this information credibly. In particular, DNA diagnostics promises to offer an order of magnitude more information about health risks than in the past, and it will not be hard to share such information credibly – perhaps using labs chosen by insurers.

The fact that individual information can be acquired by, and credibly shared with, would-be insurers ameliorates the asymmetric information and adverse selection problem. This reopens the possibility that insurance can be provided efficiently through the market. For each group with members who have identical risk profiles there would now be a separate insurance plan with its own cost and replacement rate. Each of these would simply be a point on the 45-degree line in Figure 1. More realistically, we can allow for some modest risk-heterogeneity within groups that is unknown to the insurer (or to the insured, for that matter). As long as insurers have enough information to distinguish members of different groups, we would get a series of distinct ("pooled") equilibria/plans as illustrated in Figure 2 (here we have assumed there are individuals in the entire 0-1 interval).

In this brave new world of near-complete information there would be a well-functioning market for the "creampuffs" – those individuals with low risks that insurance companies crave. In fact, anyone below the mean ($\overline{p} = .5$ in Figure 2) would be better off in such a world, assuming (as before) that private provision is no more or less efficient than public provision. This is because everyone with below-average risk subsidizes those with above-average risk in a public system. Those with the highest risk, who also tend to have the lowest incomes, may be unable to afford private insurance.[19] We have indicated such non-insurability with the dotted line in Figure 2.

[19] The link between risk and low income is thoroughly documented in the medical literature (Chen and Miller 2013; Neckerman and Torche 2007; Pampel, Krueger, and Denney 2010; Wilkinson and Pickett 2009).

Figure 2: Equilibria in the private insurance network game with shared information



Note that the possibility of credibly sharing private information has implications even for those who want to protect their privacy. The reason is that "refusers" will be placed in a high-risk group with high premiums, and everyone in that group with risk below the group average has an incentive to divulge their information to reduce their premia. If they leave, the same is true in the remaining group, and this process will continue until all the "lemons" have been called out. The logic is similar to Akerlof's except that the result will be a segmented private market for risk, with an uninsurable group relying on the (likely underfunded) public system.

Would there be majority support for privatization? The answer is most likely yes. Rehm (2016) shows that the distribution of risk is virtually always right-skewed – at least in the case of unemployment – so if people and insurers are fully informed, the median in the distribution will prefer a private system.[20] In an up-down vote, self-interested voters would therefore support privatization. Again, this conclusion is only true in a world of complete and shared information,

[20] That risk distributions are skewed to the right is also the case for subjective risk perceptions. The Survey of Economic Risk Perceptions and Insecurity (Hacker, Rehm, and Schlesinger 2013) contains survey items that elicit subjective probabilities regarding the following risks: job loss; inability to work due to serious illness; serious medical out of pocket expenditures; losing one's partner; needing to financially help out a family member; losing one's home. All of the distributions of subjective risk perceptions are right-skewed (Rehm 2016: Figure 2.6).

and we are not yet in such a world. At the same time we are almost certainly moving towards it, and as we do, the public system will become increasingly contested and fragile.

3.4. The role of government partisanship

We have already shown how preference for the level and domain of social insurance is divided by class, conceptualized as a combination of income and risk, and that such division may have risen with the information revolution. But partisanship has many sources. The strength of unions affect the ability of left parties to mobilize support (Huber and Stephens 2001; Korpi 1983; Stephens 1979), cross-cutting cleavages may undermine the support for redistribution and the left (Manow 2015; Roemer 1999; Shayo 2009; Van Kersbergen and Manow 2009), and PR electoral systems are more likely to produce center-left governments (Austen-Smith 2000; Iversen and Soskice 2006). To the extent that government partisanship is exogenous to information for these reasons, how would it affect the evolution of private markets?

Our conjecture is very simple. Since left parties tend to represent low-wage, high-risk groups, while the opposite is true for right parties, we expect left-leaning governments to try to inhibit the development of private insurance markets and shore up support for the public system. There are two main mechanisms at work. One is a simple crowding-out effect. By promoting public spending on social insurance and requiring people to pay into it, the scope for a private market to emerge is diminished. Private insurers can offer attractive plans to individuals, but since people opting out of the public system may face a double payment problem – itself politically decided, of course – there is little demand. The second is to impose non-discrimination restrictions on the private insurance industry. Private insurers may have the necessary information to offer differentiated plans, but they may not be permitted to "price-discriminate" (as it is known in the industry).

3.5. Summary

Our argument is summarized in Table 1. There are two dimensions of information: (i) the level of individual information, and (ii) whether it can be credibly shared with insurers. If private information cannot be credibly shared, markets will be inefficient and a significant number of low-risk types would forgo private plans. This raises the cost of insurance for all others, and a majority would ordinarily have an interest in a public system. If individuals cannot easily observe their own risks (low information) there will be broad support for a highly risk-redistributive public system. We call this the solidaristic welfare state outcome.

When information can be credibly shared with insurers the efficiency and feasibility of private markets rises. Credible information sharing is usually accompanied by more information, and this opens the possibility of a segmented private insurance market where each risk groups gets its own plan (the size of each pool will depend on the detail of information available as well as the economies of scale by insuring more people under the same plan). When such tailored private insurance is feasible, and information is high, those with risks below the mean will prefer private provision. With a right-skewed distribution of risk this will be a majority.

Broadly speaking, the information revolution implies a shift towards market provision, as indicated by the arrow in Table 1. Yet such a shift may be halted by left governments using measures such as anti "price discrimination" regulations, or it may be blocked by the difficulty of insuring correlated risks. When markets are blocked for either reason, increased information will cause polarization in preferences over the level of public provision and the distribution of costs. This is the contested welfare state outcome.

Table 1: Information and social insurance

		Credible information sharing?	
		No (Asymmetric information)	Yes (Symmetric information)
Level of information	Low	Solidaristic welfare state	[Pooled private insurance]
	High	Contested welfare state	Segmented private insurance

4. Empirics

We offer two preliminary tests of our model. First, we explore whether improved access to medical information has promoted the growth of private life insurance markets. Life insurance is an area with only spotty existing coverage by the state, while information asymmetries have also hampered private markets. It is governed by many of the same problems of market-failure as healthcare and offers a window into this major policy arena. Our hypothesis is that *as* *information improves, and as such information can be credibly shared with insurers, private markets for life insurance will expand*. A subsidiary claim is that this expansion is slowed in countries with strong left governments.

Second, we explore the hypothesis that *improved individual information will cause a polarization in preferences over the level of public provision*, even when the underlying structure of risk does not change. Unemployment insurance is one of the oldest responsibilities of the welfare state and it offers a good test of this proposition since private alternatives are blocked by the highly correlated nature of unemployment risks. Precisely because markets are not feasible, even with better information, we can separate conflict over the level of public provision from conflict over the domain of provision (public or private).

4.1. Measurement

Beginning with unemployment, the US General Social Survey (GSS) has regularly asked people about their subjective assessment of this risk (Smith et al. 2014), and we have "objective" information about the rate of unemployment in their occupation. By measuring the correlation between subjective and objective unemployment risk we should be able to account for the degree of polarization in preferences for unemployment insurance – after controlling for dispersion in "actual" occupational unemployment risks.

To our knowledge, the GSS is the only dataset that contains a subjective unemployment risk item that has been asked repeatedly over several decades. In particular, the GSS contains the following survey item – included in 22 survey years since 1977 – which is our measure of "subjective risk":

> Thinking about the next 12 months, how likely do you think it is that you will lose your job or be laid off – very likely, fairly likely, not too likely, or not at all likely? (Answer categories [reversed]: 1 Not likely, 2 Not too likely, 3 Fairly likely, 4 Very likely).

The GSS also includes information on respondents' occupations, which we are able to standardize across all years.[21] We use Current Population Survey (Flood et al. 2015) data to derive occupational unemployment rates (OURs) for each year that we can merge into the GSS, based on respondents' occupation. These OURs are our measure of "objective risk" (Rehm 2009). Although occupational unemployment rates do not apply to any individual – who therefore cannot use them directly – we can expect subjective risk assessments to be correlated with OUR depending on the level of information. Our proxy for "information" is therefore

[21] See Appendix C for details.

calculated as the correlation between the subjective and objective unemployment risk variables, with higher correlations indicating better information.[22]

One drawback of the GSS, for our purposes, is that the survey items most closely tapping into attitudes toward unemployment insurance policies are not available very frequently, making it impossible to track developments over long time periods. We therefore use the following survey item to proxy these attitudes:

> I'd like to talk with you about issues some people tell us are important. Please look at [this card]. Some people think that the government in Washington should do everything possible to improve the standard of living of all poor Americans; they are at Point 1 on this card. Other people think it is not the government's responsibility, and that each person should take care of himself; they are at Point 5. [.] Where would you place yourself on this scale, or haven't you have up your mind on this? [Reversed scale: 5 Govt. action; 3 Agree with both; 1 People help selves]

The underlying assumption is that unemployment can lead to poverty, and that greater concern for unemployment therefore also leads to greater concern for reducing poverty. Indeed, the correlations between the above item and survey items tapping more directly into unemployment insurance attitudes are high.[23]

The GSS allows us to trace the development of information over the last 4 decades or so in the US and to explore whether better information on unemployment risk is associated with more contestation about government intervention.

To explore the effect of data on private market penetration, we turn to health information and the development of life insurance markets. Apart from modest programs for survivor's pensions, the public system offers no life insurance. This is therefore an obvious area of potential private

---

[22] For this to be an unbiased measure of information we have to impose some constraints. First define individual *i*'s subjective unemployment risk (U_i^s) as the sum of *i*'s true risk (U_i^T) and an error that is inversely proportional to information (e_i): $U_i^s = U_i^T + e_i = U_i^{occ} + \Delta U_i^{occ} + e_i$, where the true risk has been decomposed into the unemployment rate in *i*'s occupation, U_i^{occ}, and ΔU_i^{occ}, which is the individual deviance from the that rate. The correlation between U_i^s and U_i^T, $r_{U_i^s, U_i^T}$, is 1 if $e_i = 0$, and approximates 0 if the error is very large. Hence, as $r_{U_i^s, U_i^T}$ rises it implies that information improves (that e_i falls). We do not know U_i^T, but we do know U_i^{occ} and can estimate $r_{U_i^s, U_i^{occ}}$. Since $r_{U_i^{occ}, \Delta U_i^{occ}} = 0$ (by definition), if $r_{U_i^s, \Delta U_i^{occ}}$ is assumed constant then: $r_{U_i^s, U_i^{occ}}(t+1) > r_{U_i^s, U_i^{occ}}(t) \Rightarrow e_i(t+1) < e_i(t)$, and information rises. In other words, as long as the correlation between within occupational risk and subjective risk does not change (it will normally be positive), a rise in the correlation between subjective and occupational risk will imply more information.

[23] See Appendix C for details.

expansion as more medical information becomes available that can be credibly shared. Except for major epidemics, which are exceedingly rare in advanced countries, individual risks are also largely uncorrelated.

We rely on OECD data to measure the prevalence of life insurance (OECD 2015).[24] In particular, we capture private life insurance market penetration as the ratio of direct gross life insurance premiums to Gross Domestic Product (GDP) – "which represents the relative importance of the insurance industry in the domestic economy" (OECD 2015, 16).

To measure information we use the fact that health risks are closely related to the accurate diagnosis of illnesses, as well as the possibilities for treating these. This is the kind of information that insurance companies need in order to assess risk and to price policies to particular individuals or groups. Again, privacy laws prevent insurers from direct access to medical records, but individuals are free to share them.

To estimate the amount of medical information available to individuals we turn to WHO data on mortality by cause. The WHO provides what it calls a "Potential Years of Life Lost" (PYLL) estimate for each major cause of death (cancer, cardiovascular deceases, AIDS, accidents, and so on). PYLL is the difference between how long people diagnosed with a particular disease actually live and the average life expectancy (weighting deaths occurring at younger ages more heavily). Assuming that health information (in the form of accurate test results and diagnoses) is necessary for effective treatment, more effective treatment (lower PYLL) is a sufficient condition for individual information being available. So a *reduction* in PYLL is also a sufficient condition for an *increase* in information, which is the key condition for our purposes.

A low PYLL is not a *necessary* condition for the availability of information because some diseases, especially soon after being discovered, are not treatable even if they can be accurately diagnosed (AIDS is an example). This means that PYLL can be high even when information is also high. For this reason PYLL is a noisy measure of information. Note however that such noise should bias our estimates of the effects of information on market penetration downwards.

Finally, to explore whether left governments slow the expansion of private markets, we use Huber and Stephens' (2001) cumulative measure of left seats in government (divided by the number of years), starting in 1960. Since partisanship only has an effect through slow-moving regulatory measures, such a cumulative measure seems appropriate. But because the measure does not change much, and since we estimate fixed effects models, there is little scope for it to

[24] See Appendix D for details on the data and its sources.

affect market penetration directly. Instead we focus on whether left partisanship slows the progression of markets in response to information, as hypothesized above. This is done using an interaction between our measure of information and market penetration.

4.2. Information and private market penetration: the health domain

In this section, we explore the relationship between information and the prevalence of life insurance. We expect that better information regarding health risks leads to larger life insurance markets. Our measure of information is the inverse of "Potential Years of Life Lost" (PYLL) due to illness, and we standardize it to range from 0 to 1. Again, we measure a country's private life insurance prevalence by the ratio of direct gross life insurance premiums to Gross Domestic Product ("life insurance penetration"). We have an unbalanced cross-section time-series data-set.

As a first cut, we simply regress life insurance penetration on information (model (1) in Table 2); on information and cumulative left government (model (2)); and on information, cumulative left government and their interaction (model (3)) using OLS. Model (4) adds GDP growth as a control variable. The results are in line with expectations: life insurance penetration is higher in country-years with higher information (lower PYLL) and lower in countries with more prevalence of left governments. Model (3), which includes an interaction between information and cumulative left partisanship, shows that the impact of partisanship is particularly pronounced at high levels of information. This can directly be see in Figure 3, which shows predicted values on life insurance penetration based on different combinations of information and partisanship.

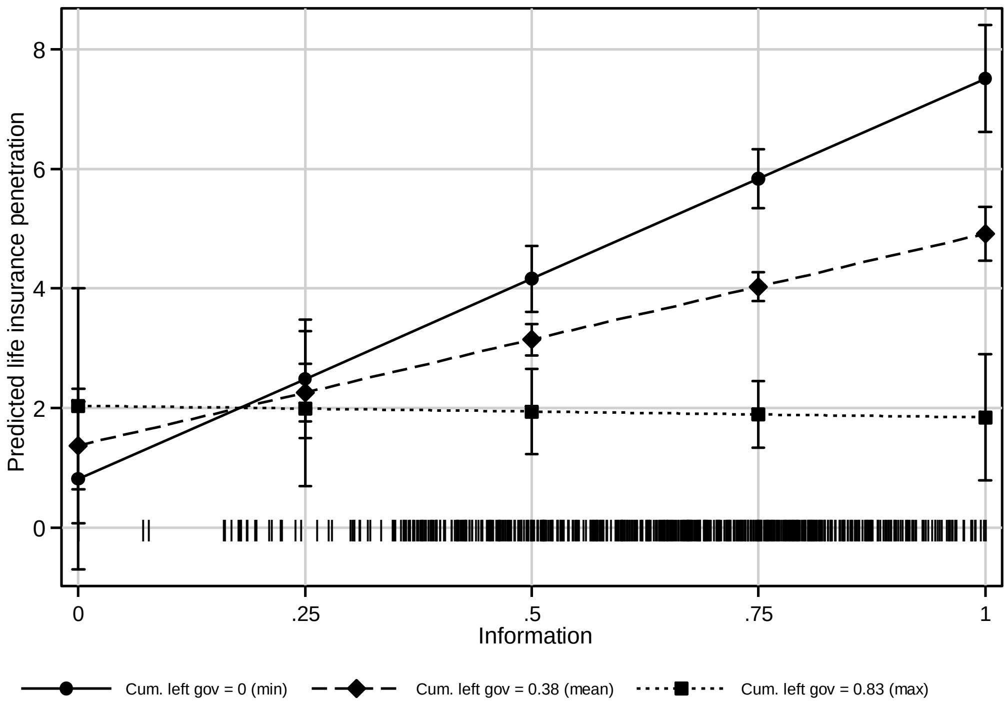
Table 2: Life insurance penetration, information, and partisanship (OLS)

	(1)	**(2)**	**(3)**	**(4)**
	Life insurance penetration (gross premiums/GDP)			
Information (- PYLL)	3.45** (0.58)	3.77** (0.56)	6.70** (1.13)	6.93** (1.14)
Cumulative left government		-4.07** (0.53)	1.48 (1.93)	1.59 (1.93)
Information X Cumulative left government			-8.31** (2.78)	-8.47** (2.78)
GDP growth				0.06 (0.04)
Constant	1.39** (0.39)	2.71** (0.41)	0.77 (0.77)	0.48 (0.79)
N	572	572	572	572
adj_r2	0.058	0.144	0.156	0.158
Log likelihood	-1379	-1351	-1347	-1346

Note: + p<0.10, * p<0.05, ** p<0.01. Coefficients displayed over standard errors in parentheses.
Estimates include control variable for break in series (not shown).

Figure 3 shows that, as hypothesized, partisanship mediates the impact of information. The estimated effects are substantively quite large. For example, life insurance penetration (premiums/GDP) is predicted to triple from slightly below 2% to slightly below 8% in a country dominated by non-left governments, simulating a change from minimum to maximum information. In countries with strong left governments there is no effect.

Figure 3: Predicted life insurance penetration as a function of information, partisanship, and their interaction



Note:
Based on model (3) in Table 2.
The rug indicates the distribution of the information variable.

These results are highly consistent with our theoretical framework's predictions. Alas, they are also based on very simple pooled OLS estimations that cannot account for the dynamic aspects of our TSCS data. We therefore estimate similar models, but use an Error Correction specification where the dependent variable is the change (first difference), not the level, of life insurance penetration (Table 3). The models include country fixed effects, and we estimate panel corrected standard errors. This test is rather demanding, but it fits the nature of our data-set much better.

The results are qualitatively comparable to the naïve OLS results presented in Table 2. The coefficients on regressors in ECMs 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 and Katz 1995). We find that there are short-term transitory effects (not statistically significant), but that the main effects are long-term and permanent. Specifically, Model (2) indicates that going from the lowest to the highest level of information raises market penetration by an average of about 3 percent (0.85+2.18) in the first year, and by 4.25 percent in the long run (0.85/-(-0.20)), which is slightly greater than the effect predicted by the simple OLS model (3.7 percent). Model (3) shows that the effect is much stronger in countries with frequent right governments, whereas it is muted in countries with frequent left governments (the pattern is nearly identical to Figure 1).

Table 3: Life insurance penetration, information, and partisanship (ECM)

	(1)	(2)	(3)	(4)
	First difference of life insurance penetration (gross premiums/GDP)			
Life insurance penetration (lag)	-0.19**	-0.19**	-0.23**	-0.22**
	(0.07)	(0.07)	(0.07)	(0.07)
Information (lag)	0.83*	0.83*	2.57**	2.75**
	(0.34)	(0.34)	(0.80)	(0.82)
Information (delta)	2.16	2.19	5.58	5.94
	(1.54)	(1.54)	(4.86)	(4.82)
Cumulative left government		0.66		
		(0.59)		
Information (lag) X Cumulative left government			-3.84**	-3.97**
			(1.43)	(1.44)
Information (delta) X Cumulative left government			-8.37	-9.23
			(10.23)	(10.16)
GDP growth				0.04*
				(0.02)
Country fixed effects	yes	yes	yes	yes
Constant	0.51*	0.26	-0.29	-0.47
	(0.21)	(0.28)	(0.33)	(0.34)
N	534	534	534	534
N of countries	22	22	22	22
R2	0.109	0.111	0.130	0.140

Note:
Coefficients displayed over panel corrected standard errors in parentheses (AR1 autocorrelation structure).
+ $p<0.10$, * $p<0.05$, ** $p<0.01$
Estimates include control variable for break in series, as well as country fixed effects (not shown).

In sum, our exploration of the relationship between health information and private insurance prevalence reveals robust evidence that they are connected as expected: better information is associated with larger private insurance markets. The estimated substantive effects are quite large – and we have good reasons to believe that they are biased downward (because our measure of information is noisy). They are also smaller in countries with left governments, arguably because left parties oppose private markets in social insurance.

As in any observational study one can raise concerns about causality, and we see the results as suggestive rather than conclusive. Yet, unlike health insurance, it seems fairly safe to rule out reverse causality since the underlying determinants of life expectancy (driving our information measure) are hardly affected by life insurance.[25] The most plausible confounder is increases in income, which could drive up demand while simultaneously increase research that might reduce PYLL (and hence our measure of information). However, when we include GDP per capita as a control, the results are virtually unaltered. Indeed, we tried a large number of potential confounders, including overall government spending, share of people over 65, and average life expectancy. None affects the reported results in any meaningful way. Improved information about individual life expectancy (based on accurate diagnostics) appears to enable insurance companies to offer life insurance plans targeted to specific risk groups, whereas in the past such plans would only attract high-risk types (lemons).

4.3. Information and polarization: the unemployment domain

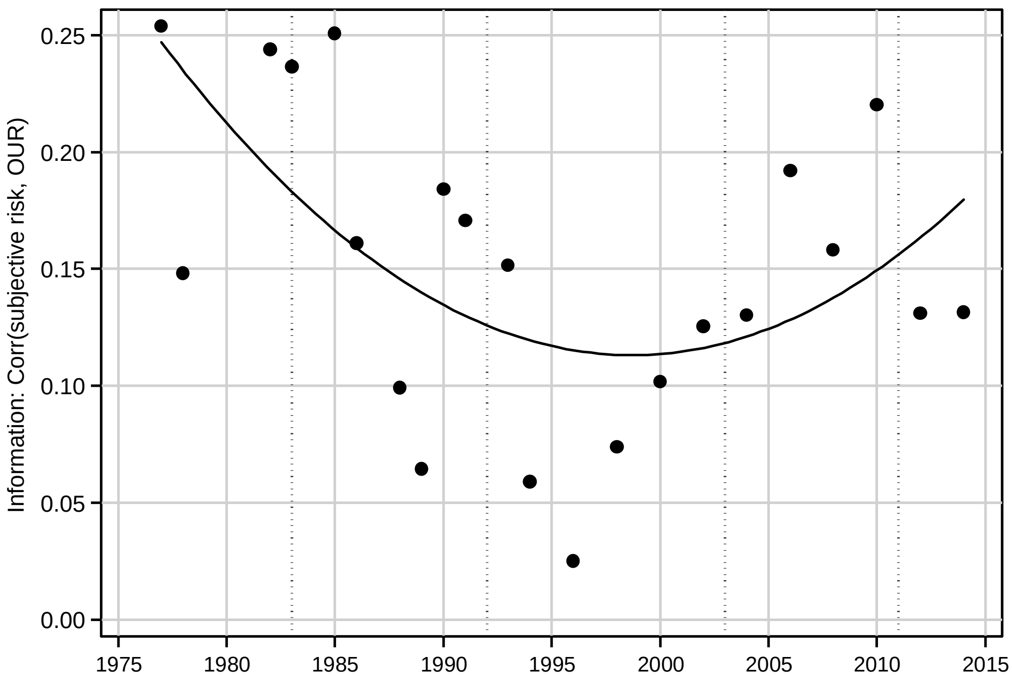
There are good reasons to believe that information in the labor market domain has been trending upward and will continue to rise. First, labor market developments increasingly sort workers into either good or bad jobs – while the middle-ground is hollowing out (Autor, Katz, and Kearney 2006; Oesch 2013; Rehm 2011a; Wright and Dwyer 2003) – with little mobility between them. With such a bipolar distribution, and little mobility, predicting one's future becomes easier. Second, those who seek information will nowadays be able to find it more easily on the internet. Labor offices around the world now provide detailed information about the current state and the predicted future of labor markets.[26] However, both of these developments are relatively recent. Another, older, trend may well have reduced information in the labor market domain: the transition from industrial to post-industrial economies (Iversen 2001). While in full swing, deindustrialization possibly made predicting one's labor market career quite difficult.

[25] The only conceivable mechanism would be moral hazard (people becoming careless with their health), but if such an effect existed it would reduce the correlation.

[26] For instance, the American Bureau of Labor Statistics (BLS) provides rich information on labor markets broken down by region, demographics, occupations, or industries – making it very easy to look up one's OUR, for example. The BLS's website also identifies fastest growing occupations, career outlooks, and the like.

Ultimately, how information in the labor market developed is an empirical question. Figure 4 provides part of the answer to that question, for the US. It plots our measure of information – the correlation between subjective and objective unemployment risk – over time, along with a quadratic fit line. The data reveal a clear U-pattern: information declined between 1975 to about 1990, and it has increased since then.

Figure 4: Development of information in labor market domain (US)



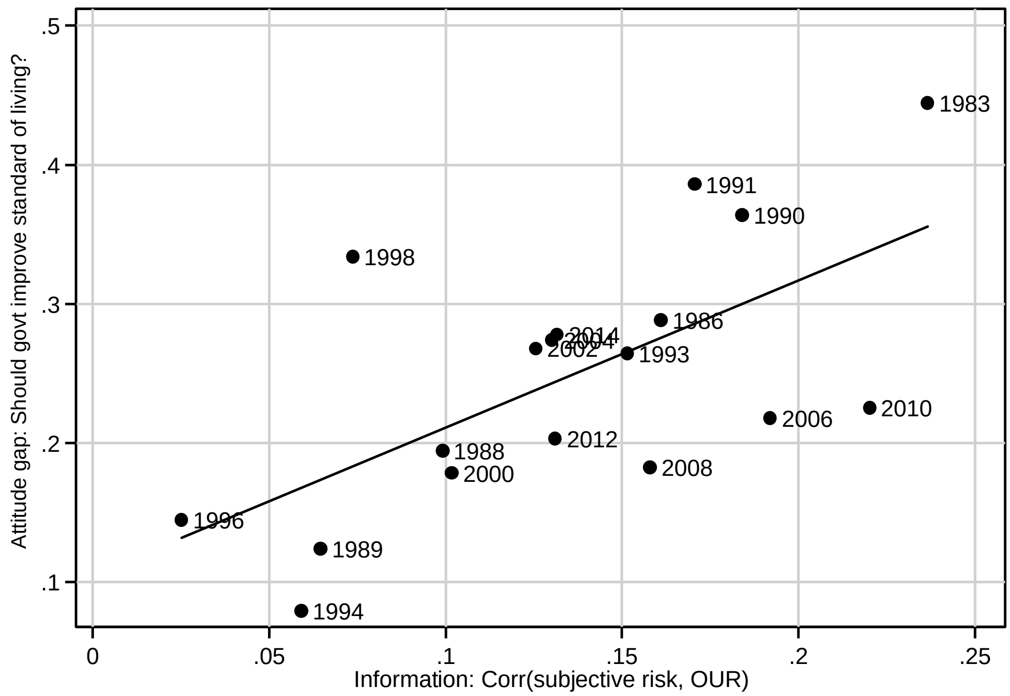
Vertical dotted lines indicate break in occupational classification

Our theoretical framework predicts that better information in the labor market domain goes hand in hand with more contestation on social policy issues. Figure 5 simply scatterplots the gap between above- and below-average risk respondents with respect to attitudes toward government intervention, as a measure of polarization, against our measure of information. It also shows the linear fit line, as well as the results from a regression of the gap on information and overall unemployment (see the note of the figure). As can be seen, there is a clear, positive, statistically significant correlation between information and contestation, as expected from our theoretical framework.[27] This relationship holds if we regress polarization on information, while controlling

[27] Interestingly, the attitudinal gap between below- and above-average risk respondents is largely – though not exclusively – driven by the increasing demand for government intervention among higher-risk respondents. While this is speculative, we would expect that support for government intervention among lower-risk respondents will

for the observed unemployment gap between the two groups. If we are right about our prediction that information in the labor market domain will continue to rise, Figure 5 provides a glimpse into the future – a future that will be characterized by a clear trend of attitudinal polarization in preferences over social insurance.

Figure 5: Information and attitude polarization (US)



Coef=1.32, SE=.38, t=3.5, adj. R2=.39. N=18 (controlling for UR)

5. Conclusion

The availability of information has rapidly increased from just a few years ago, and so has the ability to credibly share it. Most people in rich democracies today carry a smartphone with them at all times – a device that can track and transmit ones exact location, and (paired with other gadgets) ones driving behavior, step count, heart rate, blood pressure, sleeping patterns, among many other possibilities. People can now also have their entire genome sequenced for less than a

erode as information keeps rising. Our measure of information shows that the relationship between subjective and objective risk is still quite noisy, but our results show that an increase in information is, even at relatively modest levels of information, associated with an increasing gap in support for government intervention. See Appendix B.

1000 dollars. This brave new world opens up new markets for private insurers. It also has the potential to transform welfare state politics, and raises a range of new research questions.

In terms of welfare states politics, the fundamental challenge Big Data presents is that it may provide good risks with an attractive private alternative to social insurance – an alternative that was historically not viable. This opens the specter of a social insurance death spiral: since insurance is a network game, the exit of good risks makes exiting an attractive proposition for somewhat worse risks, which makes exiting an attractive proposition for even worse risks, and so on, until only bad risks are left in the (formerly common) pool. But information can have consequences even when markets are not viable. Because they scale premiums to income, not risk, welfare states are good for bad risks and bad for good risks. In contrast, private markets are good for good risks, and bad for bad risks. More information can therefore be expected to raise welfare state contestation and to increase opposition to social insurance and demand for retrenchment.

These trends are likely to be accompanied by counter-pressures and policy innovations. Regulation of information and the use of information, such as privacy laws and anti price-discrimination legislation, and well as the regulation of risky behavior itself, such as bans of trans fats or "gulp-size" soda, etc., could become hot political issues. Closely related, policies aimed at reducing risk inequality could emerge – ranging from retraining obsolescent skills, to healthcare rationing, or mandatory prenatal diagnostics. New majorities in favor of social insurance also cannot be ruled out. Changes in the distribution of risk, information, or both could be such that the median voter is a high-risk type (for example, if obesity rates keep rising). Such "risk-flips" will undermine the attractiveness of private insurance in the middle classes and shore up the public system.

We close by briefly outlining two broad research frontiers. First, the logic we have proposed potentially applies to many policy areas we have not considered here. It most obviously applies to every social insurance area, including health, disability, long-term care insurance, unemployment, and even defined-benefit pensions and other transfers that are independent of income. It may also apply to credit markets where the state plays a significant role, such as student loans, and in areas where credit is used as an alternative to traditional insurance such as unemployment insurance. Even quasi-private institutions that also play a role in social protection, notably collective wage-setting systems, are likely to be affected. Solidaristic wage bargaining in particular may depend on incomplete information because compression of wages serves as a form of insurance against downward mobility. Second, we need more information about information. We have offered some evidence on the development of private information and how it shapes social policy conflict, but our proxies are rough. Observing private information, and how citizens acquire and use it, is – by definition – a very hard problem. But thanks to Big Data, this task soon will become a lot easier.

Bibliography

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

Alt, James E., and Torben Iversen. forthcoming. "Inequality, Labor Market Segmentation, and Preferences for Redistribution." *American Journal of Political Science*.

Ansell, Ben. 2014. "The Political Economy of Ownership: Housing Markets and the Welfare State." *American Political Science Review* 108(2): 383–402.

Austen-Smith, David. 2000. "Redistributing Income under Proportional Representation." *Journal of Political Economy* 108(6): 1235–69.

Autor, David, Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the US Labor Market." *NBER Working Paper Series* (11986).

Barr, Nicholas. 2012. *Economics of the Welfare State*. 5th ed. 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." In *Handbook of Income Distribution*, eds. Anthony Atkinson and François Bourguignon. Elsevier, 679–789.

Cameron, David R. 1978. "The Expansion of the Public Economy: A Comparative Analysis." *American Political Science Review* 72(4): 1243–61.

Chen, Edith, and Gregory E. Miller. 2013. "Socioeconomic Status and Health: Mediating and Moderating Factors." *Annual Review of Clinical Psychology* 9(1): 723–49.

Cusack, Thomas, Torben Iversen, and Philipp Rehm. 2006. "Risks at Work: The Demand and Supply Sides of Government Redistribution." *Oxford Review Economic Policy* 22(3): 365–89.

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

Flood, Sarah, Miriam King, Steven Ruggles, and John Robert Warren. 2015. *Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [Machine-Readable Database]*. Minneapolis, MN: University of Minnesota [https://cps.ipums.org/cps].

Garrett, Geoffrey. 1998. *Partisan Politics in the Global Economy*. Cambridge, MA: Cambridge University Press.

Hacker, Jacob S., Philipp Rehm, and Mark Schlesinger. 2013. "The Insecure American: Economic Experiences and Policy Attitudes amid the Great Recession." *Perspectives on Politics* 11(1): 23–49.

Huber, Evelyne, and John D. Stephens. 2001. "Welfare State and Production Regimes in the Era of Retrenchment." In *The New Politics of the Welfare State*, ed. Paul Pierson. Oxford: Oxford University Press, 107–45.

Iversen, Torben. 2001. “The Dynamics of Welfare State Expansion: Trade Openness, Deindustrialization and Partisan Politics.” In *The New Politics of the Welfare State*, ed. Paul Pierson. Oxford: Oxford University Press, 45–79.

Iversen, Torben, and Thomas Cusack. 2000. “The Causes of Welfare State Expansion. Deindustrialization or Globalization?” *World Politics* 52(April): 313–49.

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

———. 2006. “Electoral Institutions, Parties and the Politics of Class: Why Some Democracies Distribute More than Others.” *American Political Science Review* 100(2): 165–81.

———. 2015. “Information, Inequality, and Mass Polarization Ideology in Advanced Democracies.” *Comparative Political Studies* 48(13): 1781–1813.

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

Manow, Philip. 2015. “Workers, Farmers and Catholicism: A History of Political Class Coalitions and the South-European Welfare State Regime.” *Journal of European Social Policy* 25(1): 32–49.

Mares, Isabela. 2003. *The Politics of Social Risk. Business and Welfare State Development*. Cambridge, MA: Cambridge University Press.

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

Neckerman, Kathryn M., and Florencia Torche. 2007. “Inequality: Causes and Consequences.” *Annual Review of Sociology* 33(1): 335–57.

OECD. 2015. *OECD Insurance Statistics 2014*. Paris: OECD Publishing. http://www.oecd-ilibrary.org/finance-and-investment/oecd-insurance-statistics-2014_ins_stats-2014-en (July 31, 2015).

Oesch, Daniel. 2013. *Occupational Change in Europe: How Technology and Education Transform the Job Structure*. Oxford: Oxford University Press.

Pampel, Fred C., Patrick M. Krueger, and Justin T. Denney. 2010. “Socioeconomic Disparities in Health Behaviors.” *Annual Review of Sociology* 36: 349–70.

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

Rehm, Philipp. 2009. “Risks and Redistribution: An Individual-Level Analysis.” *Comparative Political Studies* 42(7): 855–81.

———. 2011a. “Risk Inequality and the Polarized American Electorate.” *British Journal of Political Science* 41(2): 363–387.

———. 2011b. “Social Policy by Popular Demand.” *World Politics* 63(2): 271–99.

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

Rodrik, Dani. 1998. “Why Do More Open Economies Have Bigger Governments?” *Journal of Political Economy* 106(5): 997–1032.

Roemer, John E. 1999. “Why the Poor Do Not Expropriate the Rich: An Old Argument in New Garb.” *Journal of Public Economics* 70(3): 399–424.

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.

Rueda, David. 2007. *Social Democracy Inside Out: Partisanship and Labor Market Policy in Advanced Industrialized Democracies*. Oxford: Oxford University Press.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. Norton.

Shayo, Moses. 2009. “A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution.” *American Political Science Review* 103(2): 147–74.

Shojania, KG, EC Burton, KM McDonald, and L Goldman. 2003. “Changes in Rates of Autopsy-Detected Diagnostic Errors over Time: A Systematic Review.” *JAMA* 289(21): 2849–56.

Smith, Tom W., Peter V. Marsden, Michael Hout, and Jibum Kim. 2014. *General Social Surveys, 1972-2014 [Machine-Readable Data File]*. Chicago: NORC at the University of Chicago (Data accessed from the GSS Data Explorer website).

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

Stiglitz, Joseph E. 1982. “Self-Selection and Pareto Efficient Taxation.” *Journal of Public Economics* 17(2): 213–240.

Sweeney, Latanya. 2002. “K-ANONYMITY: A MODEL FOR PROTECTING PRIVACY.” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(5): 557–70.

Van Kersbergen, Kees, and Philip Manow. 2009. *Religion, Class Coalitions, and Welfare States*. Cambridge, MA: Cambridge University Press.

Wilkinson, Richard, and Kate Pickett. 2009. *The Spirit Level: Why Greater Equality Makes Societies Stronger*. Bloomsbury Press.

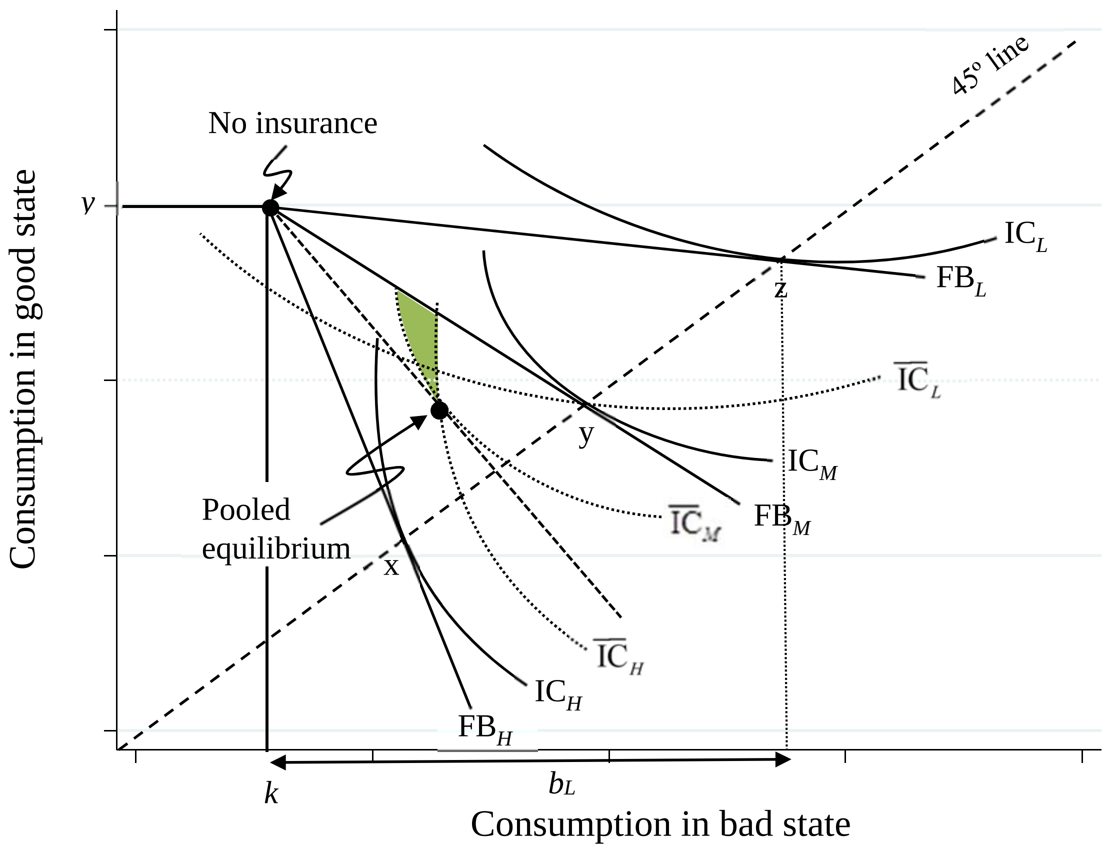
Wren, Anne, and Philipp Rehm. 2014. “The End of the Consensus? Labour Market Developments and the Politics of Retrenchment.” *Socio-Economic Review* 12(2): 409–35.

Wright, Erik Olin, and Rachel Dwyer. 2003. “The Patterns of Job Expansions in the United States, a Comparison of the 1960s and 1990s.” *Socio-Economic Review* 1(3): 289–325.

Appendix A: Graphical representatio of the pooling equilibrium with private information and adverse selection

In Figure A1 there are three risk groups, *L*, *M*, and *H*. With no insurance, income is *y* in the good state and *k* in the bad (*y* and *k* are assumed here to be the same for all groups). The solid downward-sloping "fair-bet" (FB) lines are the feasible sets of allocations of income between the two states that would have the same expected value (in the model, the slopes of these lines are $-(1-p_i)/p_i$). Each risk groups would want to allocate enough income for insurance to equalize income in the two states, which are where the fair-bet lines intersect the 45-degree line and the indifference curves (IC) are tangent to each FB line. If risk was common knowledge, and insurance markets competitive, each group would be offered a contract corresponding to these points, denoted *x*, *y*, and *z* in the figure (assuming no costs of provision and zero profits). The benefit, *b*, stipulated in each contract is the difference between income in the good and bad state, which is $b_i = \dfrac{\pi_i \cdot y_i \cdot (1-p_i)}{p_i}$ in our model (illustrated here along the x-axis in the case of *L*).

Figure A6: Example of a pooled equilibrium with three risk groups



Yet, insurers cannot observe the risks of different groups, and instead have to pool these so that the expected combined payouts are equal to total insurance payments. When risks are pooled across all three groups, the lines are drawn so that this expected payout (contract) line is equal to *M*'s fair-bet line. Imagine now that the insurer offered *y* as an insurance plan. If all bought the plan the insurer would break even (the zero-profit condition would hold) and the outcome would be sustainable from the perspective of insurers. But *L* would not buy this plan since *L*'s indifference curve is below the "no insurance" point: *L* would be worse off with insurance than without. With *L* opting out there is a new pooled fair-bet/contract line between *M* and *H*, which is the downward-sloping dashed line. A pooled equilibrium is now feasible since any point on that line (above the 45-degree line) is superior to the no insurance point for both *M* and *H*.

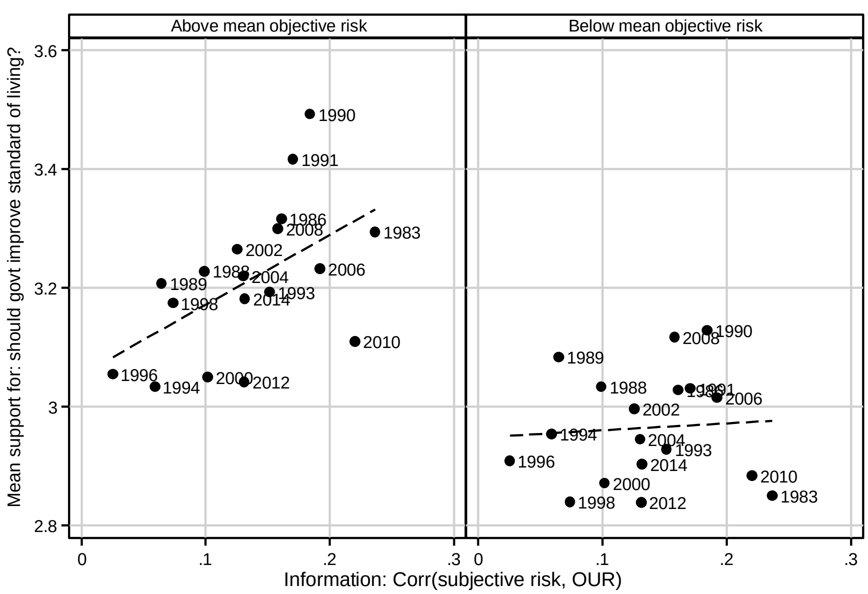
Note that there is a shaded area above the pooled equilibrium point where *M* would be better off (while *H* would be worse off). In the Rothschild-Stiglitz model a competitive firm could move into this space and make a profit by selling to *M* (the lower risk group) only. That would undermine the pooled equilibrium. But this assumes that insurance companies can monitor the quantity of insurance bought by different groups since otherwise they might end up selling to a high-risk type for a loss. If insurers cannot not do this, or if they cannot distinguish between low-risk types with high risk-aversion and high-risk types with low risk-aversion, the pooled equilibrium is feasible (but inefficient). This would also be true if insurers can "collude" to not move away from the pooling outcome, and regulators should not object to this as it preserves a competitive market. In our model pooling equilibria are allowed for one or all these reasons.

Finally, the pooled equilibrium is represented as a single point on the contract line, whereas in the model we have allowed different points on the line in proportion to how much people pay into the system. In that case, the point in Figure A1 represents an average.

Appendix B: Information and polarization

Figure 4 in the main text shows that attitude polarization – the gap between above- and below-average risk respondents with respect to attitudes toward government intervention – positively correlates with our measure of information. Polarization in response to information could be due to low-risk respondents moving away from high-risk respondents, the reverse, or both. Figure A2 shows that the effect is driven by increased support for interventionist policies among high-risk workers. This is not entirely surprising when recalling that there are two steps involved in increasing individual information: the availability of observable signals and the acquisition of information. We suspect that low-risk workers are better informed because they tend to have higher education and be in stronger social networks where information is shared (Iversen and Soskice 2015). High risk-types, on the other hand, depend more on receiving strong signals from their environment in order to update their risk perceptions, and such signals have intensified as the risk structure has become more bifurcated.[28] The absence of information effects at the low-risk end may thus indicate that these workers are already informed, but this is an area for future research.

Figure A7: Information and attitude polarization II (US)



Left panel: Coef=1.97, SE=.45, t=4.3, adj. R2=.51. N=18 (controlling for UR)

[28] But note again that the relationship between our measure of information and preference polarization persists even after controlling for the gap in "objective" unemployment rates, so there is an independent effect of information itself.

Appendix C: Information and attitude polarization

Data

This analysis is primarily based on the General Social Survey (GSS) 1972-2014 cumulative file. See gss.norc.org.

Details on the subjective risk survey item

Question wording: "Thinking about the next 12 months, how likely do you think it is that you will lose your job or be laid off – very likely, fairly likely, not too likely, or not at all likely?" (Answer categories [reversed]: 1 Not likely, 2 Not too likely, 3 Fairly likely, 4 Very likely).

This is variable "joblose" in the GSS 1972-2014 cumulative file, available for the following years: 1977, 1978, 1982, 1983, 1985, 1986, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, and 2014.

Details on occupational variables

We use the following occupational variables in the GSS:

- occ [Rs census occupation code (1970)]: 1972-1987
- occ80 [Rs census occupation code (1980)]: 1988-2010
- occ10 [Rs census occupation code (2010)]: 2012-2014

To standardize these different occupational classifications over time, we exploit the fact that the Current Population Survey (CPS) not only codes occupations into the relevant SOC classification for a given year (SOC1970, SOC1980, etc.) but also into SOC1990 for all years. This allows us to use the CPS data to back out a concordance between the SOC-classifications for various years and SOC1990. We use these concordances to code the GSS occupational data into SOC1990.

We use Current Population Survey (Flood et al. 2015) data to derive occupational unemployment rates (OURs) for each year, coded in the same SOC1990 classification, and merge them into the GSS, based on respondents' occupation. To make sure that results are not influenced by outliers, we top-code OURs at 25% (roughly p99).

Attitudes toward unemployment insurance policies

We proxy attitudes toward unemployment insurance policies with the following survey item: "I'd like to talk with you about issues some people tell us are important. Please look at [this card]. Some people think that the government in Washington should do everything possible to improve the standard of living of all poor Americans; they are at Point 1 on this card. Other people think it is not the government's responsibility, and that each person should take care of himself; they are at Point 5. [.] Where would you place yourself on this scale, or haven't you have up your mind on this?" [Reversed scale: 5 Govt. action; 3 Agree with both; 1 People help selves]

This is variable "helppoor" in the GSS 1972-2014 cumulative file, available for the following years: 1975, 1983, 1984, 1986, 1987, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, and 2014.

The GSS contains two survey items that capture attitudes toward unemployment insurance more directly:

- survey item 'aidunemp' ("On the whole, do you think it should or should not be the government's responsibility to F. Provide a decent standard of living for the unemployed")
- survey item 'jobsall' ("On the whole, do you think it should or should not be the government's responsibility to A. Provide a job for everyone who wants one").

These items overlap with the 'helppoor' item we use in the analysis for the years 1989, 1990, 1996 ('aidunemp') and 1989, 1990, 1991, 1996, 1998 ('jobsall'), respectively, and they correlate with that item at 0.3147 and 0.3347 (statistically significant at $p<0.01$).

Cross-national data

Our analysis is based on American data. For other OECD countries we have data going back to 1990, and we find a robust upward trend in the correlation between subjective and objective unemployment risk (our measure of information) in the cross-national data, mirroring the upward trend in the US since the 1990s in Figure 4.

The cross-national data are based on a variety of surveys, namely:

- European Working Conditions Survey (EWCS)
- European Quality of Life Survey (EQLS)
- European Social Survey (ESS)
- International Social Survey Programme (ISSP)
- Eurobarometer (EB).

Appendix D: Information and life insurance penetration

Life insurance penetration data

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

Here are summary statistics for the life insurance penetration variable in the sample: Mean: 3.64, SD: 2.79, Min: 0.04, Max: 20.02, P1: 0.08, P5: 0.23, P10: 0.81, P25: 1.84, P50: 3.06, P75: 4.99, P90: 6.38, P99: 16.19.

Mortality by cause

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). The original data source is the WHO data Mortality Database (http://www.who.int/healthinfo/mortality_data/en/).

We make use of the "Potential Years of Life Lost" (PYLL) variable, defined in the following way: "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 standardised, for each country and each year. The total OECD population in 2010 is taken as the reference population for age standardisation. 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 (Accessed on 01 September 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 behavioural 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% 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 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.

<u>Sample</u>

The following country-years are in our sample, which is determined by data availability: AUS (1984-2011, N=24), AUT (1987-2013, N=27), BEL (1983-2012, N=30), CAN (1984-2011, N=28), CHE (1983-2012, N=30), DEU (1990-2013, N=21), DNK (1983-2012, N=28), ESP (1983-2013, N=31), FIN (1983-2013, N=31), FRA (1983-2011, N=29), GBR (1996-2013, N=17), GRC (1992-2012, N=20), IRL (1983-2010, N=27), ISL (1983-2009, N=26), ITA (1983-2012, N=28), JPN (1983-2013, N=31), NLD (1995-2013, N=19), NOR (1983-2013, N=30), NZL (1989-2003, N=10), PRT (1983-2013, N=28), SWE (1983-2013, N=30), USA (1983-2010, N=27).