Should I Use That Rating Factor?
A Philosophical Approach to an Old Problem

Prepared by Chris Dolman, Seth Lazar, Tiberio Caetano and Dimitri Semenovich

Presented to the Actuaries Institute
20/20 All-Actuaries *Virtual* Summit
3-28 August 2020

Institute of Actuaries of Australia
ABN 69 000 423 656
Level 2, 50 Carrington Street, Sydney NSW Australia 2000
t +61 (0) 2 9239 6100 **f** +61 (0) 2 9239 6170
e actuaries@actuaries.asn.au **w** www.actuaries.asn.au

Abstract

A common question asked of insurance professionals is whether or not a rating factor *ought* to be permitted for price setting. To answer this question one often falls back on a set of heuristics, for example, is the factor commonly used in the market, does it predict well, would people generally expect it to be used, or is it visible to customers. However, the underlying principles behind such intuitions are not necessarily appreciated or understood.

In this paper, we draw on contemporary ideas in procedural and distributive justice to decompose the classical notion of risk into risk types that in turn allow development of semi-formal criteria by which rating factors may be evaluated. We hope that this new set of conceptual tools will allow practitioners to reason with greater clarity about ethical questions surrounding pricing systems.

Keywords: pricing, ethics, rating, data

Motivation

Should I use that rating factor? This is a question regularly posed by, and asked of, insurance professionals. In our experience, there are three common situations which give rise to this form of question:

1. *New external data becomes available*
   A new dataset originating from outside the organisation becomes available to use. This situation has been long predicted to become more common in the contemporary "big data" era (see Actuaries Institute (2016) for an extended discussion). Notably, this is increasingly likely to include a large number of potential additional rating factors rather than a single factor. This situation often arises from a discussion between a data vendor and an insurer.
2. *New rating questions are proposed*
   A proposal is put forward to collect additional data directly from customers, through adding a question to a policy application form. During the process of exploring whether to ask for this additional information, the question of whether the factor ought to be used is naturally asked. We note that recent trends in data portability - for example the Australian Consumer Data Right (CDR) - could enable more data to be requested from consumers without the burden of manual data entry that may have prevented this historically.
3. *Retrospective review of practices*
   In the process of a pricing review a question arises whether certain rating factors in use today ought to be permitted, perhaps with the original rationale for the decision to use the factor no longer readily available. This prompts an investigation seeking to understand and perhaps challenge the current position.

However, in each of these cases, the meaning of the question is slightly different:

1. In the case of new external data, the primary area to explore is usually the origins of the data and its perceived legitimacy for pricing, which is likely a secondary use of that data and often invisible to the consumer. Data about people, especially personally identifiable information, is usually more concerning, and more likely to lead to serious questions, than data about assets like houses or cars. Legal questions are often posed - though if this is a vendor proposal these are usually adequately addressed. The connection of the data in question to risk may or may not be intuitive - if not this may be an additional line of inquiry.
2. For a new rating question the considerations are typically quite different. Transparency is innate to the proposal (though perhaps degraded somewhat in a CDR era, particularly in the detail), and some form of perceived legitimacy for use of the data in pricing is usually required due to this level of transparency. What remains are considerations such as whether being asked the question is seen as

socially acceptable, and whether the customer is able to answer the question with suitable accuracy.

3. A retrospective review may re-examine longstanding answers to questions outlined above. Since such questions are commonly answered using judgement and heuristics, a new perspective may result in a different answer. Over time, social norms may also change - what was deemed acceptable in the past may no longer be so.

The main observation to make is that in all of the cases above, the question: "Should I use that rating factor" is imprecise. Hence, our first task is to refine the question so it admits a more formal analysis.

The refined question: "Should I use that data to incrementally improve my knowledge of risk and then change prices in direct proportion to the incremental knowledge gained?"

As observed above, "Should I use that rating factor?" is not a question of sufficient specificity to allow formal analysis.

To add clarity, we may consider the situation in which the question is posed: the setting of the price of an insurance contract. Since an insurance contract is one of risk transfer, and since risk clearly does vary between individuals, it follows that this difference in risk might be a legitimate reason for price differences between contracts. Simply: the more the risk being transferred, the larger the price.

Under this definition of "risk", we are focusing on the expected value of claims costs only, ignoring any differences in the variance, kurtosis or other characteristics of the distribution of claims cost which might exist across the population, though we note that these might also be included under some other definitions of "risk".

The typical reason for wishing to add rating factors to the existing set of rating data is some belief that the additional data will allow incremental improvement in the accuracy of the prediction of the expected cost of claims under the contract (typically known as the "risk premium" or "pure premium"). In the remainder of this paper, we will refer to prices set directly in line with statistical estimates of risk as "risk prices", "pure premium" or "risk premium".

Of course, there is no guarantee that the rating data is solely used in a manner that reflects some prediction of risk. It is common for insurers to add other considerations into the rate setting process, notably to allow for expenses, profits and to account for market dynamics such as competitiveness. However, since the primary motivation for adding data is usually to improve risk prediction, for simplicity we will restrict our analysis and assume that rating data is ultimately used to set prices directly in accordance with estimates of the risk premium.

With this restriction of scope, we can refine the question being asked: "Should I use that data to incrementally improve my knowledge of risk, and then change prices in direct proportion to the incremental knowledge gained?" Using this refined question, we may now explore the ways in which a factor relates to risk and any estimate of the pure premium, and whether acting on such relationships is one which ought to be permitted.

In considering how much information ought to be used, there are two extreme cases along a continuum: the situation of no information at all about risk, and the situation of unlimited or complete information. Traditionally, insurance professionals have tended to argue that more information is always a good thing, allowing finer and finer grained estimation of risk at the individual level. This paper serves to challenge that traditional mindset, providing a framework by which we can determine how close to the "perfect information" state we ought to get. Or as Cass Sunstein puts it in a recent book, we sometimes need to understand what we *don't* want to know (Sunstein 2020). We return to these extreme limiting states at the end of the paper.

Recent Related Work

This topic has been studied extensively over the years, both in academia and professional circles. Often the discussion focuses on discrimination in a legal sense, though concepts may often be generalised away from this strict legal framing. Recent work from Frees and Huang (2020) gives a thorough overview of the historic literature and global regulation. Loi and Christen (2019) analyse the merits, drawbacks and tradeoffs of predictive accuracy and statistical discrimination in the context of insurance pricing. Popular debate often inspires extensive analysis of a single item of data, for example the recent global debate around genetic testing data (Bélisle-Pipon et al (2019), Doble & Chen (2017), Otlowski et al (2019)) and consideration of sex around the time of the Test-Achats ruling in the EU (Schmeiser et al (2014), ABI (2010), Michael et al (2012)). Some researchers have focussed on the downstream question of how to avoid discrimination based on a factor, even indirectly - Lindholm, Richman, Tsanakas & Wuthrich (2020) recently provide such a formulation.

Our framing of the general question - around knowledge of risk - distinguishes our analysis from many other works on this topic. By first separating analysis of the information being sought (that of risk) from the representation of that information (i.e. data), we can provide a systematic way to think about the right actions to take which goes beyond considerations of data.

Preliminary Step - "Could I Use That Rating Factor?"

Before considering whether we *should* use a rating factor, we should first establish whether we in fact *can* use that factor at all.

There are usually legal restrictions surrounding use of data, which should first be considered. Notably, anti-discrimination legislation in many countries will prohibit the use of certain facts about people within many decision-making contexts, including insurance pricing. Privacy law may also prohibit the collection or use of certain items of data. There may also be contractual restrictions on data use - for externally acquired datasets in particular, this should be carefully evaluated.

The data should be checked for accuracy, completeness and reasonableness. Clearly, data which is unreliable in such a manner is not necessarily fit for purpose for an important economic decision like insurance pricing. Unreliability might be remedied by allowing the consumer to correct any errors in the data.

Finally, we should validate that the data does in fact relate to risk in some manner. This might involve testing the data within an existing modelling framework for significance. If the data does not in fact relate to risk, then we have no need to use it.

Once we have established that the data *could* be used for risk pricing, we can proceed to considering whether it actually *should* be. The remainder of this paper outlines a framework by which this form of normative decision can be made, going beyond the baseline requirements of society as described under the law.

Epistemic Reasons to Reject a Factor - Evidence of the Wrong sort of Risk

Is it always reasonable to set prices in line with as precise as possible an estimate of the risk premium as we can create? Traditionally, insurers have answered such questions in the affirmative. If we answer "yes", then adding more knowledge about risk into the pricing decision ought to always be permitted, and the answer to our question is trivial and uninteresting.

Modern theories of justice would challenge this answer. On some (especially libertarian, e.g. Nozick, 1974) theories of justice, the costs of all bad outcomes should 'stay where they fall'. If something bad happens to you, that's your business alone. But on other theories, the costs of some bad outcomes that befall a person should be fairly shared by others, according to some principle of distributive justice (e.g. Dworkin, 1981). In the insurance context, "bad outcomes" means claims under an insurance contract. We can naturally extend outcome-based concepts of justice such as those above to the ex-ante probabilities of bad outcomes. This then corresponds to a decision as to when insurance premiums should follow risk, or some other premium allocation mechanism.

With this in mind, we can decompose the total risk into two categories:

1. those elements of risk for which premiums ought to be determined in line with estimates of the risk premium
2. those elements of risk for which premiums ought not to be determined in line with estimates of the risk premium, but rather through some principle of distributive justice

For any new rating factor to be considered, we propose we must ultimately determine whether it provides evidence of type-1 risk or of type-2 risk. If it is evidence of type-2 risk, it means we have *epistemic* grounds not to use the rating factor: it gives us evidence of the wrong kind of risk.

One common way to delineate between type-1 and type-2 risks is to consider responsibility for the risk. If you can be reasonably held responsible for a risk you are exposed to, this is likely to represent a type-1 risk, otherwise it is likely to be type-2.

There are likely to be a range of other considerations for delineating risks into type-1 and type-2, and it is likely that people will disagree on the process of delineation. This is to be expected - as noted above there are varying theories of justice which will each, if followed, give different specifications. Users of this framework will need to impose their own preferred method of delineation in order to apply it. Where people disagree on outcomes, we consider it more productive for practitioners to debate the underlying concepts of risk and justice imposed by them to generate their delineation, than to begin by debating data.

Below, we identify several categories of risk that we consider plausibly to be type-2, and hence for which it might not be considered just to include into the calculation of the risk premium.

Innate risks

The most widely accepted distinction between type-1 and type-2 risks focuses on whether it is possible for the individual to control whether or not the bad outcome occurs. If it is within their control, it is likely to be a type-1 risk. If it is not, then it is probably a type-2 risk, and should not (in principle) affect their pure premium.

Risks caused by factors innate to oneself are clearly uncontrollable. No matter what action one takes, the risks (and any bad outcomes arising from those risks) are not able to be ameliorated by those actions. These bad outcomes may be said to be a result of *constitutive luck* (Nagel, 1979).

Genetic diseases are a good example of this. We cannot control our genetics. But some bad outcomes are caused by them. Genetic information that provides evidence that a person is likely to suffer down the line, because of their genetics, is evidence of type-2 risk, and plausibly should not be used to calculate pure premiums.

Similarly, some bad outcomes might be causally linked to other innate factors that are not within your control. Race, for example, might be a predictor of mortality, but since your race is not in your control, anti-discriminaton legislation prevents this evidence being legally used to set your premium. There are some notable exceptions to anti-discrimination legislation for insurance companies, some of the reasons for which we explore briefly below.

Although some theories of justice take the view that ‘I am not my brother’s keeper’, most find some way to socialise the costs of bad luck of which we are not in control. Some defend distributing those costs equally to all (Cohen, 1989); others allow faultless harm to remain where it falls, provided the person bearing it remains above some threshold of wellbeing (Casal, 2007); while others eschew both thresholds and equality, and simply place greater weight on helping those who are worse off to begin with (Parfit, 2000). Regardless of precisely how we redistribute these risks, most theories of justice would reject the idea of using them to determine an individual’s pure premium. They are evidence of the wrong kinds of risks.

Essential, or Required, Risks

Some risks are literally out of our control - there is nothing we can do to ameliorate them. Other risks are controllable in principle, but only if we take on excessive costs. For example, if you drive a car then, no matter how cautious a driver you are, there is some risk that your car will malfunction and cause significant harm, to yourself or others. The only way to eliminate this risk is not to drive at all. But in many countries and contexts it is unreasonably costly not to be able to drive. So while in one sense this risk is controllable, in another it is not. It may therefore count as a type-2 risk, which should be accommodated through some distributive principle—for example, a Compulsory Third Party (CTP) insurance scheme with a fixed pricing schedule.

Some instances of strict liability may fall into this category. Society may determine that an individual should be held liable for an outcome even in the absence of fault or poor intent on their part. For example, consideration of compensation for "blameless accidents" has formed part of some CTP schemes in the past. In these cases, society has decided to assign liability to an individual (to allow a claim for CTP), because society has deemed this to be an appropriate mechanism for creating just outcomes for those injured. But perhaps the cost of it should not fall there, and should be otherwise distributed.

In some cases, the costs of control may vary significantly across the population, leading to different conclusions depending on the circumstances of the individual considered. For example, driving in the middle of the night is generally found to be more dangerous than driving during the day, and for many trips the choice of when to make it may be easily controlled by the individual. One can decide when one goes to the supermarket. With this in mind, we might say that risks caused by driving at night represent a type-1 risk. However, for a shift worker, their job may require travel at night and so it may be unreasonable to ask this individual to change their behaviour and control this risk.

Some risks are controllable in principle, and controlling them does not put unacceptable costs on the individual, but instead has significant social costs. For example, working or volunteering as a firefighter significantly increases one's risk of personal injury or illness. One can lower that risk easily enough: don't work or volunteer as a firefighter. So in that sense, those who choose to fight fires are responsible for the additional risks that they face. For most people this risk-reducing decision is not unduly costly. But if everyone took this approach, then we wouldn't have any firefighters. The risk is individually controllable, but socially uncontrollable, in the sense that it comes at too great a social cost. So at least some of the risks associated with working or volunteering as a firefighter are type-2 risks, which should then be redistributed according to some principle of distributive justice (Eyal, 2006).

Risks Imposed by Others

Some risks to an individual are created by the action of another, but still borne by the individual. For example, a victim of domestic abuse may be more likely to be affected by poor health and claim on their health insurance, but this risk is outside of their control and imposed upon them, so ought not to be factored into their pure premium. In many cases these risks will fall into one of the categories just considered: they will be strictly out of the individual's control, or else within their control, but too costly to avoid. But sometimes neither of these will be true. It might be possible, even beneficial, to leave the abuser. It might be predictable that they will abuse. And yet even if the victim does not leave the abuser, it still would not be appropriate to hold them to account for the harms they predictably suffer, when they are brought about through the wrongdoing of another. In our view, these are also type-2 risks.

Application

This is not an exhaustive list of type-2 risks, and some may disagree that those listed above should be type-2. When considering whether we can use a rating factor to determine an individual's pure premium, we should examine whether it is evidence of type-2 risk. If it is considered type-2, then it shouldn't be used in line with an estimate of risk premium, but rather filtered through some suitable principle of distributive justice, and at least partially socialised (which principle one should adopt is a matter for a separate discussion).

An interesting challenge arises when we cannot tell whether a rating factor provides evidence of type-1 or type-2 risk. For example, big data analytics often provide statistical evidence *that some outcome will likely occur*, without providing evidence of what will cause the outcome (Pearl, 2018). Without that causal information, it may be impossible to tell - from that data - whether the outcome is within the individual's control or not.

For example, suppose we have access to a huge dataset of social media and internet browsing data, which enables us to infer that some individual is at a significantly higher risk of personal injury over the next five years. We may not be able to extract from the data the reason *why* this person is at higher risk. But this matters a lot. Suppose they are in fact at greater risk because they have an abusive partner, which is being indirectly revealed in their social media and browser activity. Then, as we noted above, that's evidence of a type-2 risk, the costs of which should be (more or less) socialised. But suppose it's because of some hobby that they have freely chosen, in full knowledge of the dangers to themselves. Then it is a type-1 risk, and they should rightly bear its costs.

When we do not know whether a given rating factor provides evidence of type-1 or type-2 risk, clearly the first thing to do is to try to find it out. New techniques bringing causal reasoning into machine-learning-based data analytics (e.g. Pearl, 2009; Morgan & Winship, 2014), as well as enabling counterfactual testing of classification and regression algorithms, would be relevant here. However the current level of maturity of these tools often does not allow the drawing of conclusive causal inferences.

Morally speaking, we should consider the costs of accidentally classifying a risk as type-2, when it is in fact type-1, and vice versa, in order to act appropriately under this uncertainty.

An important subtlety in insurance contracts is the size of contingent payment or choice of the "sum insured". Generally, we consider this to be controllable, and a type-1 risk - even where the underlying risk of a bad outcome is not controllable. Extending one of the examples above, whilst it is not reasonable to use race to set life insurance premiums (since it is innate), it would be reasonably expected that a choice to buy a term life insurance policy for twice the sum insured should come at twice the price, even if that choice were generally taken by one race rather than another. This pricing structure would be an example of indirect discrimination, but one which we believe is defensible, being reasonable in the circumstances.

Social and Moral Reasons to Reject a Factor

Even if we have shown that a given rating factor is a legitimate candidate for type-1 risk, there may be other considerations which could apply. Sometimes even the right kind of evidence ought not be considered, because there are social or moral reasons not to do so. The situation is analogous to one in the criminal law, where evidence that might be relevant to determining a defendant's guilt - for example, their record of having committed similar crimes - might be excluded for substantive moral reasons (these are subject to debate, but one reason to exclude evidence of prior convictions is a commitment to giving people a genuine second chance).

Below, we outline several plausible social or moral reasons to reject a rating factor. Again, this list is by no means exhaustive.

The data, or inferences from the data, may be seen as too private or personal

The most obvious example of a social or moral reason not to attend to some prospectively informative rating factor is that doing so would excessively undermine individual privacy. Return to the example of using non-traditional big data to inform premium pricing. Would you be comfortable knowing that your browsing history, everything you do on social media, and all data generated from your smart devices, were being used to determine your insurance premium? They might well provide valid evidence of some type-1 risks - indeed, at sufficient scale they almost certainly would do. And yet, many people in Western liberal democracies with strong privacy regimes and, in some cases, rights to privacy, would deny that it is appropriate to use such personal information in setting insurance premiums. In some cultures this may be different - the most obvious example being China, where detailed surveillance of individuals is generally more accepted.

One possible response is to give people the chance to opt in to providing their personal information, in order to reduce their premium. We already have evidence that many people are willing to make just this trade. But if the total risk of the population stays constant, then there is no way to reduce the premiums of those who opt in, without in the long run raising those of the people who opt out. The consenting behaviour of some creates pressure on others, to which they do not consent. What's more, if enough people voluntarily share their data, their doing so will enable strong inferences to be drawn about others regardless of whether they consent to share their data or not (Crawford and Schultz, 2014). This is a notable issue for genetic testing data in some countries, where some moratoriums on insurance use of genetic testing data allow the declaration of positive test data (i.e. proof of low-risk) at the customer's option (for Australia, see FSC (2019), for the UK, see ABI (2018)). Left unchanged, this current policy position seems likely to create a problem in the long run, as more and more members of the public take genetic tests. As has been shown, time and again, in the age of big data individual consent is only modest protection for individual and social privacy (Barocas and Nissenbaum, 2014).

Data may be intuitively irrelevant

The rating factor in question may be epistemically legitimate, but may still fail a general test of human intuition. If the typical insurance customer does not intuitively believe that the data ought to relate to risk in the way it does, reality may be of little comfort to them. To be more concrete - an insurer may have evidence to suggest that a risk is type-1, but the public might not accept this evidence as correct (for no clear, rational reason). This is an emotional objection which a public brand ought to take very seriously.

For example, suppose an occupation carries with it some risk of significant injury, but those risks are ameliorated by a set of safety standards for that work. An insurer may have evidence to suggest that the risks are still present to some degree, but those people in that line of work may believe the risks have been entirely removed. In deciding whether to use this evidence, the insurer should consider whether the beliefs (however unfounded) of this group should be taken into account, or risk brand damage.

The general discussion within the fair machine learning literature around communication and explanation of algorithmic decisions is clearly related to the requirement above (Selbst and Barocas, 2018). With effective communication of the reasons for a decision, it may be possible for mismatches between intuition and reality to be reduced significantly.

Using the data may result in disparate impacts on different populations

Some rating factors might provide evidence of type-1 risks, but might nonetheless result in premiums varying significantly across racial or other protected demographic groups.

Where the groups in question are morally significant, for example because they have been the historic victims of structural discrimination, there is cause for concern. Suppose it were true, for example, that members of a minority group were more likely to have particular types of vehicle accidents in which they were at fault. If that minority group is, in general, subject to structural discrimination, then there might be grounds to exclude that rating factor from your deliberations when setting the risk premium. In our view, the protections of anti-discrimination law are partly grounded in the fact that one’s membership in a protected group is often evidence of type-2 risk, or at least evidence of risk that could be type-2 or type-1, and partly in the role of the law in remediating historical discrimination.

We encourage the reader to check that their considerations here (and anything that might be inferred from our comments above) are compliant with anti-discrimination law, and to seek legal advice where they are unsure.

Using the rating factor might have unacceptable unintended consequences

The use of risk data to price insurance does not occur in a vacuum. If their premiums are significantly affected, then people will respond to these price signals, and may change their behaviour accordingly. Even if some rating factor provides evidence of type-1 risk, is not too private, is accepted by the public, and is not discriminatory, it might still be unsuitable for use in setting premiums if the unintended consequences of doing so are bad enough. For example, if people know that their genetic information will be used to change their premiums, should they discover it, this might give them reason not to, for example, take a DNA sequencing test. And yet there may be significant health and other benefits in an individual knowing their genome. So this might be another good reason not to use genetic information for setting insurance prices.

The data might relate to past behaviour for which the individual should not now be held accountable

With insurance as with criminal justice, it might sometimes be important to give people a second chance. What one did in the past might be evidence of one's type-1 risk, but might still not be appropriately used in setting one's insurance premium. For example, no doubt the fact that a person has served a custodial sentence is evidence of many different kinds of type-1 risk. But we might take the view that, having served their sentence, they should not be further 'punished' by their criminal past increasing their insurance prices. We have social/moral reasons to reject that rating factor's use, even if it is predictive of the right kind of risk. We might take a similar view in other cases: for example, historical exposure to a toxic chemical might increase your risk of a future illness, but this could be considered out of the control of the individual now and restricted for pricing a future insurance contract.

Risk Ignorance of the Insured

A common justification for risk pricing is that it operates as a "risk signal" - allowing the public to understand the level of risk by way of a simple number, and incentivise adaptive actions to reduce risks by basic economic incentives driven from this number. This effect is believed by some to improve the resilience of society over the medium to long term.

An important consideration, therefore, is whether the knowledge about risk gained from the data is gained by both the insurer and the insured, or merely by the insurer. If an insurer gains knowledge about the risk of an individual but the consumer remains ignorant (save for the overall price level), the consumer may reasonably claim that they cannot know which adaptive action to take to reduce their risk. The commonly held justification noted above is then invalid.

This is likely to be a notable issue for external data an insurer may collect without the knowledge of the consumer - for example detailed data about the built environment. It may also be an issue where very granular consumer data is provided with the consumer's knowledge and consent, but where the consumer may not be truly aware of the detail within it (for example any detailed data shared under regimes like CDR).

An important mitigating action is transparency and communication of the knowledge the insurer has about the risks of the individual consumer, particularly those items of knowledge that are very significant for the individual insured, and which may be affected by their behaviour.

Desire for Solidarity

The origins of insurance stem from a social desire for solidarity - individuals choosing to come together to share their individual risk by pooling relatively equal contributions. This led to the establishment of mutual companies and friendly societies. There may still be cases where this social desire for mutuality persists, which would give us reasons to reject using a rating factor.

For example, private health insurance in Australia operates using a "community rating" mechanism - generally, individual risk-rating is not utilised in setting premiums. Whilst some of the reasons listed previously might also drive this form of scheme design, the overarching rationale is perhaps one of solidarity. The risk of poor health is one we are all exposed to, and mutualisation of this risk is generally seen as appropriate by many in this country.

Consequential Reasons to Still Use a Factor

A traditional argument used by the insurance industry to justify the use of uncontrollable risk factors (i.e. evidence of type-2 risk) is the impact of anti-selection in non-compulsory insurance markets. Whilst anti-selection can come in many guises, in this case we are particularly concerned about the market participation effects caused by non-use of a rating factor.

For example, if age were not used to set prices for term life insurance, we would find that prices would increase dramatically for younger people, and would fall for older people. This might make policies unattractive for younger people in general, leading to falling market participation and social problems associated with un-insurance when insurable events (in this example, deaths) occur. The average costs of the remaining pool of risks would rise, leading prices to increase, causing a potential self-reinforcing effect. In the extreme, this may cause markets to collapse (Rothschild and Stiglitz 1976, Akerlof 1970).

It was with such arguments that the insurance industry traditionally argued for exemptions under certain anti-discrimination laws - notably age, sex and disability/health discrimination legislation. The social costs of not allowing the use of the rating factor were deemed to outweigh the moral costs of the discrimination. We note the legislative position is regularly debated and is by no means fixed, with perhaps the most notable recent example of change being the outlawing of sex in insurance pricing under the Test-Achats ruling of the ECJ (European Commission, 2011).

Competitive dynamics are another anti-selective consideration. If a rating factor is in general use, removal of that factor will generally come at some cost to the insurer who removes it: they will be "selected against" by competitive market forces. In such cases, there may be a need for market collaboration or agreement in order to remove a factor deemed problematic, which may also require approval of competition regulators, as such behaviour could be viewed as anti-competitive.

In general, where there are significant consequences associated with not using a rating factor, we believe it is legitimate to weigh those consequences against the costs of using the factor when considering whether it ought to be used. We note these costs will tend to be higher where a factor is in general use.

Examples of applying the framework

Credit score data for comprehensive motor insurance

Credit score data is used in many markets around the world to rate motor insurance, and is commonly found to predict the risk of accidents reasonably well.

Suppose an insurer does not currently use such data for motor insurance pricing, and suppose for now that the use of this data in this context is allowed by law (in some countries this is not the case).

Using the framework above, we must first attempt to identify whether someone's credit score is evidence of type-1 or type-2 motor insurance risk. We feel it is not obvious that it represents type-1 risk, and so should be treated with caution. Indeed, it's not immediately obvious exactly how someone's credit rating might affect the risk of a car insurance claim, even if this is found to be statistically valid.

Even if we were convinced that this represented evidence of type-1 risk, there are clear social reasons to reject the use of the factor. One's credit rating is primarily evidence of ability to repay a loan, and one could reasonably object on privacy and relevance grounds to this being used for car insurance pricing. Those with poor credit history are also more likely to be otherwise more vulnerable members of society, and penalising this group further with increased insurance premiums may not be socially acceptable.

In some markets, however, use of credit rating data for motor insurance pricing is commonplace. We suggest that if this is case for the market in question, the anti-selection risks for not using the factor may be dire enough that use of the data could still be permitted, overriding the concerns above. In such circumstances, it would appear that some industry-wide reform of practices could be beneficial to society, and the insurer might feel obligated to support such reforms.

Wearable data for health and life insurance (e.g. fitbit, smart watch)

With the growth in consumer friendly IoT devices, notably those associated with sports and activity, there is the potential for insurers to gain information relating to people's lifestyles. This may be particularly relevant for health and life insurance.

The core purposes of such devices is to incentivise healthier lifestyle choices. It might seem reasonable, therefore, to allow consumers to demonstrate their healthier choices by sharing this data with their insurer, to obtain a better price.

However, we might challenge the use of the data on grounds of accuracy. Wearable devices are fallible, and do not always measure what is actually happening (the authors can attest to this from personal experience). Users of such devices might also not wear them consistently, or might attempt to deliberately manipulate the outputs - particularly if they are to be used for something like insurance pricing.

Putting this aside for now, if we accept that this data accurately encodes lifestyle choices which lead to improved health outcomes (or the reverse), and provided we accept these choices as the responsibility of the individual, it seems reasonable to classify them as type-1.

However, not all data collected by such devices will fall into this category. Heart rate or blood pressure data, for example, may be encoding a type-2 risk (such as an underlying health condition caused by genetics). Care is needed to assure ourselves that the data truly does encode type-1 risk, before it is considered for use.

In this example we must also recognise the data as non-universal. Not everyone has a smart device tracking their habits. So we must also consider the impact of the use of the data on those without access to such data. Analysis of this impact will depend on the size of the effect measured by the data, the proportion of the population opting in to data sharing, and whether the opt-out group contains a significant proportion of otherwise disadvantaged people.

We suggest that in the present day, it is likely that some use of such data would be permitted for insurance pricing. There are some fairly simple implementations of this in-market today, often combined with a more general reward programme. For example, Discovery's "Vitality" programme generates reward points for healthy activities (some measured via wearables[i]), which give access to increasing discounts on insurance products[ii]. However, we suggest that as uptake of such devices increases, the cost of such products on those without devices will increase, so any decision to use such data should be regularly reviewed.

Topographical data for property flood insurance

With the increase in resolution and reduction in cost of aerial imagery, detailed topographical maps of areas are increasingly used by insurers for premium rating. This has led to significant price changes for those exposed to topographically driven risks, most notably floods.

Certainly, topographical data directly relates to the risk in question. In the case of private property, there is an argument that the risk is controllable at the individual level - a person can sell their property if they choose. Hence, we would generally argue that the risks are type-1.

However, whilst this may be true at the individual level (though there are certainly arguments to be had about hardship for that individual), at the societal level, insurance affordability for individual high-risk properties that are already built is a great concern. At the extreme, the risks for a property may be unaffordable for most individuals. This can cause stress on communities. Some socialisation of the costs of risks may be desirable on social grounds, assuming there is broad support for these communities to be retained. This must be balanced against the need to incentivise risk reduction, and incentivise risk avoidance for newly built properties. The design of Flood Re in the UK is an example of a mechanism which takes the latter into account, subsidising flood insurance for many existing high-risk properties but not properties built in the future[iii].

Theoretical Limiting Cases

As shown above, in real-world settings our framework may be used to help us assess whether a rating factor ought to be used. But what about in the extreme, idealised limiting cases we mentioned previously? Does our framework still hold?

The blind insurer limit: insurance under zero information

In this scenario, assume we have a *blind insurer* that has no information about customers, at the time of signing the insurance contract, that allows differentiating their risk.

In this case, we have no evidence available to differentiate risks, either type-1 or type-2. There seems to be no principle that can justify treating customers differently (at least in a systematic way). Based on this, our framework would suggest we need to set premiums in some just manner. We could either have all premiums follow the same sum insured curve (*equal premiums*) or perhaps have different premiums but would have to allocate them in some random way proportional to sum insured (*random premiums*).

Whilst unrealistic (certain details about the customer would always be known), this does demonstrate that the logic in our framework should hold as this limiting case is approached.

The oracle insurer limit: insurance under complete prior information

The other end of the spectrum is an *oracle insurer* that knows with certainty all information prior to issuing an insurance contract. We assume this allows complete and certain knowledge of the expected costs under the policy - though uncertainty still remains as to the actual future realisation of claims.

In this case, using our framework, we should ask which component of the risk is type-1 and which is type-2. We expect that the more risk we determine to be type-2, and the more we choose to distribute the costs of those type-2 risks across the population, the more unattractive this market might become for low-risk individuals, and the more prone to failure this market may be (owing to anti-selection, as described briefly above). This assumes that low-risk individuals are aware that they are low-risk, and can choose to exit the market without severe personal impact. In such cases, some form of regulated scheme, either privately or socially underwritten, may be required.

For type-1 risks, however, the situation is different. If we assume the risk information is *shared* with the insured, this provides an incentive to act to reduce these type-1 risks. This is the traditional "price signal" argument taken to the extreme - an expensive policy provides an economic incentive to reduce one's risk. In such a case, the oracle insurer may attempt to predict people's modified behaviour as a result of receiving the risk information in order to set an appropriate price for what would remain of their risk, or may choose to only issue contracts assuming current behaviour, together with regular monitoring of it (for example at each policy renewal).

Conclusion

In a world awash with data, it is becoming more and more necessary for insurance professionals to consider whether available data *ought* to be used for price setting. Traditional heuristics may have served us well for a long time, but insurance professionals are generally not familiar with the underlying logic for such heuristics, and may find it challenging to extend them to a big-data world.

In this paper, we have presented a high-level framework by which professionals can begin to reason about this question, without falling back on common heuristics.

Firstly, we consider epistemic reasons for rejecting a rating factor - whether it's evidence of the wrong kind of risk to be priced at a risk premium level.

Secondly, we identify social or moral reasons for rejecting a rating factor (even if it is epistemically useful). This amounts to a set of reasons for creating wilful ignorance.

Thirdly, we consider the consequences of rejecting a rating factor, and for cases where this is severe, weigh the costs of rejection against the costs of use.

We then illustrate the application of this framework with some case studies, and finally, we consider how the proposed framework applies to idealised limiting cases of no information and complete information for an insurer.

We encourage insurance professionals to take our framework here as a starting point when considering whether a rating factor ought to be used.

References and Notes

Actuaries Institute (2016); "The Impact of Big Data on the Future of Insurance" https://actuaries.asn.au/Library/Miscellaneous/2016/BIGDATAGPWEB.pdf

ABI (2018); Association of British Insurers and HM Government "Code on Genetic Testing and Insurance" https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/751230/code-on-genetic-testing-and-insurance.pdf

Akerlof, George A. (1970); "The Market for "Lemons": Quality Uncertainty and the Market Mechanism" The Quarterly Journal of Economics, Vol. 84, No. 3. (Aug., 1970), pp. 488-500.

Association of British Insurers (ABI) (2010) 'The Use of Gender in Insurance Pricing: Analysing the Impact of a Potential Ban on the Use of Gender as a Rating Factor, ABI Research Paper No. 24' London: Association of British Insurers

European Commission (2011); "Sex Discrimination in Insurance Contracts: Statement by European Commission Vice-President Viviane Reding, the EU's Justice Commissioner, on the European Court of Justice's ruling in the Test-Achats case" https://ec.europa.eu/commission/presscorner/detail/en/MEMO_11_123

Frees, Edward (Jed) W. and Huang, Fei, "The Discriminating (Pricing) Actuary" (May 4, 2020). Available at SSRN: https://ssrn.com/abstract=3592475 or http://dx.doi.org/10.2139/ssrn.3592475

FSC (2019) "FSC Standard No. 11 Moratorium on Genetic Tests in Life Insurance", 21 June 2019. https://www.fsc.org.au/resources-category/standard/1779-standard-11-moratorium-on-genetic-tests-in-life-insurance/file

Lindholm, Mathias and Richman, Ronald and Tsanakas, Andreas and Wuthrich, Mario V., Discrimination-Free Insurance Pricing (January 16, 2020). Available at SSRN: https://ssrn.com/abstract=3520676 or http://dx.doi.org/10.2139/ssrn.3520676

Loi, Michele and Christen, Markus; "Insurance Discrimination and Fairness in Machine Learning: An Ethical Analysis" (August 17, 2019). Available at SSRN: https://ssrn.com/abstract=3438823 or http://dx.doi.org/10.2139/ssrn.3438823

Nagel, Thomas (1979); "Mortal Questions"; New York: Cambridge University Press

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: The economics of markets with imperfect information. Quarterly Journal of Economics 90, pp. 629–649.
Barocas, S. and Nissenbaum, H. (2014), 'Big Data's End Run around Anonymity and Consent', in Julia Lane, et al. (eds.), Privacy, Big Data, and the Public Good: Frameworks for Engagement (New York: Cambridge University Press), 44-75.

Bélisle-Pipon, J., Vayena, E., Green, R.C. et al. (2019) Genetic testing, insurance discrimination and medical research: what the United States can learn from peer countries. Nat Med 25, 1198–1204. https://doi.org/10.1038/s41591-019-0534-z

Casal, Paula (2007), 'Why Sufficiency Is Not Enough', Ethics, 117 (2), 296-326.

Cohen, G. A. (1989), 'On the Currency of Egalitarian Justice', Ethics, 99 (4), 906-944.

Crawford, Kate and Schultz, Jason (2014), 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms', Boston College Law Review, 55, 93-128.

Dworkin, R (1981), 'What Is Equality? Part 2: Equality of Resources', Philosophy and Public Affairs, 10 (4), 283-245.

Doble, A and Chen, J. (2017) 'Genetics – A testing Time for Insurers?' Actuaries Institute Dialogue. https://actuaries.asn.au/Library/Miscellaneous/2017/TheDialogueWEB.pdf

Eyal, Nir (2006), 'Egalitarian Justice and Innocent Choice', Journal of Ethics and Social Philosophy, 2 (1), 1-18.

Michael, L. et al (2012) 'Fairness in Insurance Pricing'. Report of the IFoA GIRO Working Party. https://www.actuaries.org.uk/system/files/documents/pdf/e3-paper.pdf

Morgan, S., & Winship, C. (2014). Counterfactuals and Causal Inference: Methods and Principles for Social Research (2nd ed., Analytical Methods for Social Research). Cambridge: Cambridge University Press.

Nozick, R. (1974). Anarchy, state, and utopia (Vol. 5038). New York: Basic Books.

Otlowski, M.; Tiller, J; Barlow-Stewart, K; Lacaze, P (2019) 'Genetic testing and insurance in Australia'. Australian Journal of General Practice. doi: 10.31128/AJGP-10-18-4722

Parfit, Derek (2000), 'Equality or Priority', in Matthew Clayton and Andrew Williams (eds.), The Ideal of Equality (Basingstoke: Palgrave), 81-125.

Pearl, Judea. (2018). Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution. WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 3-3. 10.1145/3159652.3176182.

Pearl, J. (2009). Causality. Cambridge: Cambridge University Press.

Schmeiser, H., Störmer, T. & Wagner, J. (2014) 'Unisex Insurance Pricing: Consumers' Perception and Market Implications.' Geneva Pap Risk Insur Issues Pract 39, 322–350 (2014). https://doi.org/10.1057/gpp.2013.24

Selbst, Andrew D. and Barocas, Solon (2018), 'The Intuitive Appeal of Explainable Machines', Fordham Law Review, 87, 1085-1139.

Sunstein, C. R. (2020). Too Much Information: Understanding What You Don't Want to Know. MIT Press.

---

[i] https://www.discovery.co.za/discovery_coza/web/linked_content/pdfs/vitality/vitality_news/vitality_fitness_points_table.pdf

[ii] https://www.discovery.co.za/life-insurance/discounts-and-paybacks

[iii] https://www.floodre.co.uk/can-flood-re-help-me/eligibility-criteria/