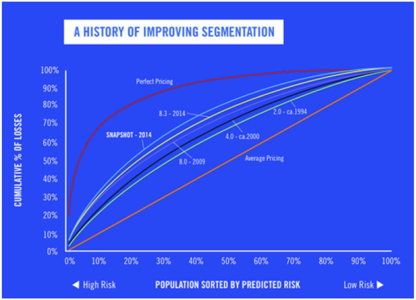
Predoc - II

Philipp Ratz

December 05, 2022

A brief History

Model Comparison and Lorenz curves



Source: Progressive Insurance

Act. Non-Vie 2017/18 at ENSAE

Table of Contents

Overview (of the past)

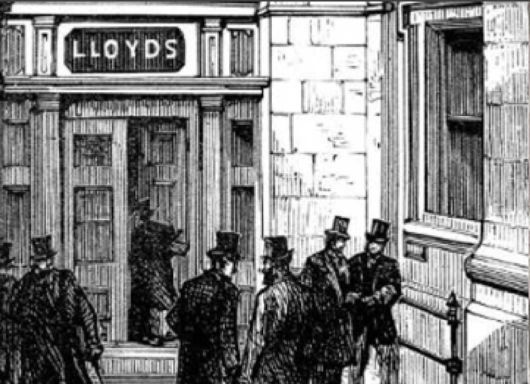
Current progression (1.25 years in)

Overview (of the present)

1. Motivation & global problem statement
2. Existing work in the field, yet-to-be done
3. Insights from research, summary from articles
4. Further research and ideas

Code can be found on phi-ra (access token required for working papers), illustrations on the blog

Risk sharing on Networks



Motivation

CMO NETWORK

The Top 5 Industries Most Hated By Customers

Blake Morgan Senior Contributor

I am a Customer Experience Futurist, Author and Keynote Speaker.

Follow

Source: [Forbes, 2018]

Insurance is among the top-5 hated industries in America according to *Forbes*. Often there is inherent **distrust** between the parties and insurance is looked at as an investment.

In a sense - peer-to-peer (P2P) insurance brings back the roots. Considering insurance less of a gamble, but more as mutual assistance.

Literature I

Important in risk sharing is the concept of conditional mean risk sharing: $g_i^*(S) = \mathbb{E}[X_i|S]$, for $S = \sum_i^n X_i$ where $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a random vector of n risks. This is studied for example in [Denuit and Robert, 2020b] or in [Denuit and Robert, 2020a].

[Abdikerimova and Feng, 2021, Feng et al., 2022] build upon this and derive specific P2P pareto optimal solutions, here specified as an optimization problem. They also introduce heterogeneity in the risk pool.

But these approaches all implicitly impose a structure on the underlying network

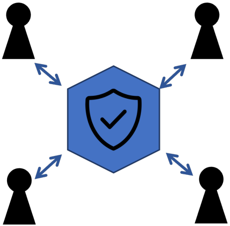
Literature II

Trust and prediction problems have been extensively analysed in P2P *lending*. For example, [Granovetter, 2005, Lin et al., 2013] find that P2P networks have a lot of information about the quality of loan allocation.

It it also shown that P2P can have complementary (rather than substituting) effects within the market [Liu et al., 2020]. But these findings are rarely used in the P2P-insurance case.

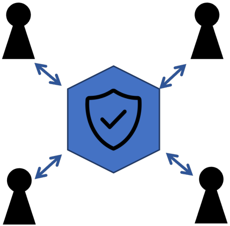
Can this be translated to the insurance case?

A different classification

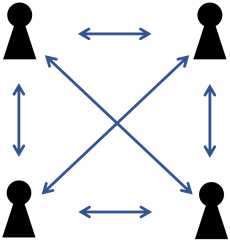


Traditional Model

A different classification

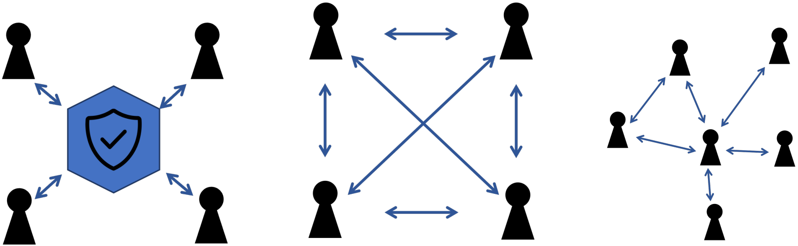


Traditional Model



P2P "Classic"

A different classification



Traditional Model P2P "Classic" P2P "Network"

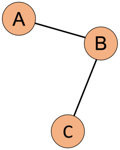
A new approach

1. Consider *arbitrary* networks
   i. Common language for networks
   ii. Simulation approaches
2. Set up basic risk-sharing scheme
   i. Evaluate using simulations
3. Improve mechanism

Presentation based on insights from *Collaborative Insurance Sustainability and Network Structure* [Charpentier et al., 2021]. Submitted and presented at conferences IME, Astin (2021), seminars for AxA and UQAM (2021) and conference CSSC, Informs (2022)

Graph

A graph $\mathcal{G}$ is an ordered pair $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents a set of vertices (or nodes) and $\mathcal{E}$ a set of edges (or links) such that $\mathcal{E} \subseteq \{\{x, y\} | x, y, \in \mathcal{V}, x \neq y\}$

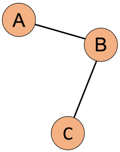


Graph

A graph $\mathcal{G}$ is an ordered pair $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents a set of vertices (or nodes) and $\mathcal{E}$ a set of edges (or links) such that $\mathcal{E} \subseteq \{\{x, y\} | x, y, \in \mathcal{V}, x \neq y\}$

Adjacency Matrix

The adjacency matrix A for a set of vertices $\mathcal{V} = \{v_1, \ldots, v_n\}$ is a matrix of size $n \times n$, such that $A_{i,j} = 1$ if there is an edge from vertex v_i to v_j.



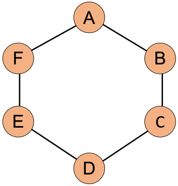
$$\begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

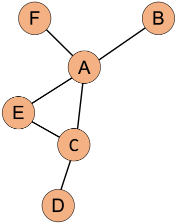
Graph Terminology II

Degree distribution

The degree of a vertex is the number of edges that are incident to it. We can construct the degree vector from elements d_i as $d = A\mathbb{1}$. The average degree of a network can be calculated as:

$$\bar{d} = \mathbb{E}[d] = \frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}} d(v) = \frac{2|\mathcal{E}|}{|\mathcal{V}|} = \frac{1}{|\mathcal{V}|}\,\|d\|_1$$



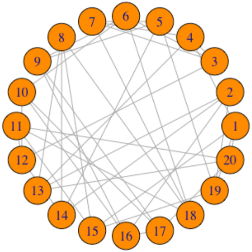


Graph Simulation - I

Regular Network

Every node in $\mathcal{V}$ has degree equal to some constant $k < n$. As a technical condition least one of k and the number of vertices must be even.

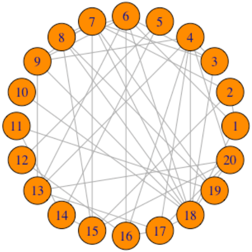
$$\mathbb{E}[d] = k$$
$$\mathbb{V}[d] = 0$$



ErdősRényi Network

Each possible edge $i, j \in \mathcal{V} \times \mathcal{V}$ is included in the network with probability p independent from other edges. The Network then has a degree distribution that follows a binomial distribution $\mathcal{B}(n-1, p)$. Also, if p not too large but n_V: $p \sim \lambda / n_V, d(v) \sim \mathcal{P}(\lambda)$

$$\mathbb{E}[d] \approx \mathbb{V}[d]$$



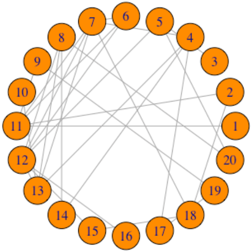
Graph Simulation - III

Preferential attachment Network

Consider starting from a small network and adding new nodes. During every step, a new node is connected to the existing nodes with a probability p that is proportional to the degree of an existing node d_i. This way, every new node is more likely to connect to existing "popular" nodes. Here the degree distribution follows a power law.

$$\mathbb{E}[d] = ?$$
$$\mathbb{V}[d] = ?$$

usually fat tailed

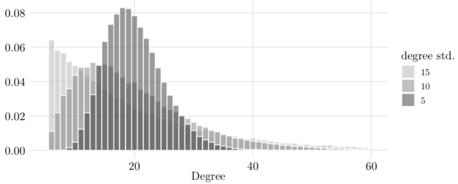


Graph Simulation - IV

Instead of the mentioned algorithms, we use a flexible method:

$$d \overset{\mathcal{L}}{=} \min\{5 + [\Delta], n - 1\}, \quad \Delta \sim \mathcal{G}(\alpha, \beta)$$

Where we vary $\frac{\alpha}{\beta^2}$ but not $\frac{\alpha}{\beta}$ (fix the mean $\overline{d}$ but change variance $\mathbb{V}[d]$)



The Core: Risk Sharing with Friends

- An agent assumes reciprocally risk from friends and shares their losses

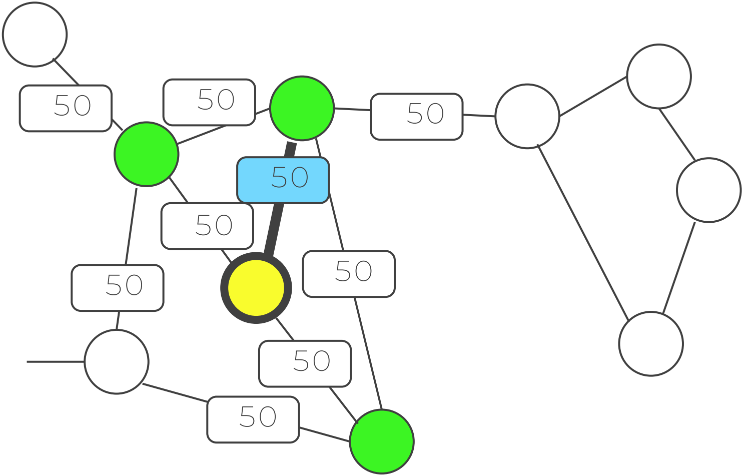
The Core: Risk Sharing with Friends

- An agent assumes reciprocally risk from friends and shares their losses
- We consider a homogenous population where a risk materialises at most once per period

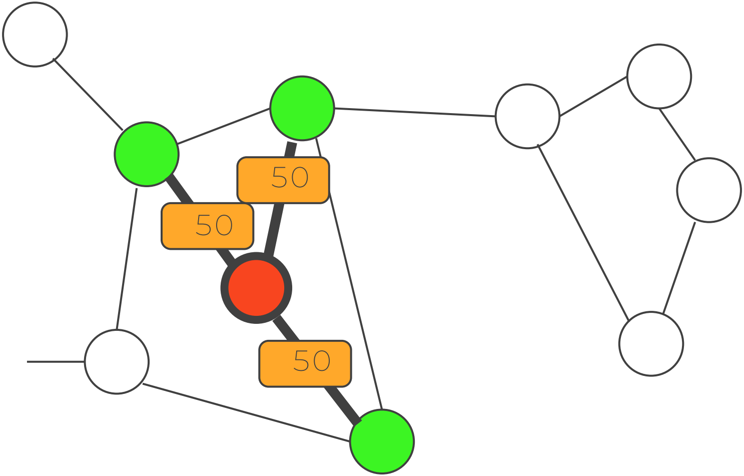
The Core: Risk Sharing with Friends

- An agent assumes reciprocally risk from friends and shares their losses
- We consider a homogenous population where a risk materialises at most once per period
- With homogenous risks usually implies the $\frac{1}{n}$-rule (and everybody gets the same contract). We refer to the contract amount as γ

Risk Sharing on a network



Risk Sharing on a network



The Core: Risk Sharing with Friends

- Here we consider the case where the homogenous population only wants to insure an amount below a deductible

Deductible:	$s_i = s \quad \forall i$
Risk (i.i.d.):	$Y_i \sim 100 + \mathcal{G}(a, b)$
Helper:	$Z_i \sim \mathcal{B}(p)$
Loss:	$L_i = Z_i \min\{s, Y_i\}$
Engagement:	$\gamma_{(i,j)} = \gamma_{(j,i)} = \dfrac{s}{d}$

Single period wealth

Let $\mathcal{V}_i$ denote the set of "friends" of a node i. We denote the wealth after one period:

$$\xi_i = Z_i \min\{s, Y_i\}$$

Single period wealth

Let $\mathcal{V}_i$ denote the set of "friends" of a node i. We denote the wealth after one period:

$$\xi_i = Z_i \min\{s, Y_i\} - Z_i \min\{d_i\gamma, \min\{s, Y_i\}\}$$

Single period wealth

Let $\mathcal{V}_i$ denote the set of "friends" of a node i. We denote the wealth after one period:

$$\xi_i = Z_i \min\{s, Y_i\} - Z_i \min\{d_i \gamma, \min\{s, Y_i\}\} + \sum_{j \in \mathcal{V}_i} Z_j \min\left\{\gamma, \frac{\min\{s, Y_j\}}{d_j}\right\}$$

Single period wealth

Let $\mathcal{V}_i$ denote the set of "friends" of a node i. We denote the wealth after one period:

$$\xi_i = Z_i \min\{s, Y_i\} - Z_i \min\{d_i\gamma, \min\{s, Y_i\}\} + \sum_{j\in\mathcal{V}_i} Z_j \min\left\{\gamma, \frac{\min\{s, Y_j\}}{d_j}\right\}$$

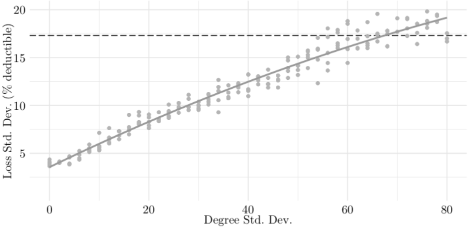
1. **Simulation:** Draw a graph, fix s, draw from $\mathcal{B}$, $\mathcal{G}$
2. **Optimization:** $\xi_i = f(s, Y_i, Z_i, d_i, \gamma)$
3. **Compare:** ξ vs. $\min\{s, Y\}$

The simple framework

Consider a network with $\overline{d} = 20$, $s = 1000$, $\mathbb{E}[L] = 4.5\% \times s$, $\sqrt{\mathbb{V}[L]} = 17\% \times s$

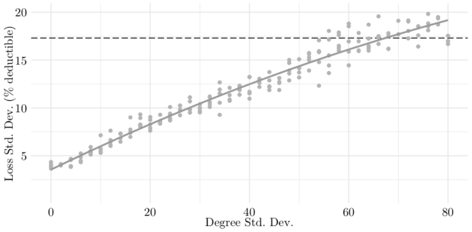
The simple framework

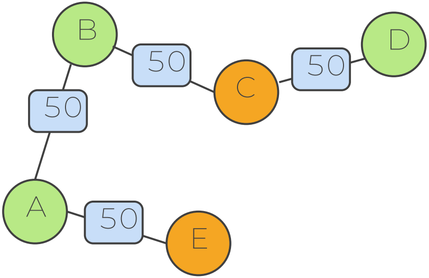
Consider a network with $\overline{d} = 20$, $s = 1000$, $\mathbb{E}[L] = 4.5\% \times s$, $\sqrt{\mathbb{V}[L]} = 17\% \times s$



The simple framework

Consider a network with $\overline{d} = 20$, $s = 1000$, $\mathbb{E}[L] = 4.5\% \times s$, $\sqrt{\mathbb{V}[L]} = 17\% \times s$





The simple framework

How did we do?

- ☑ Risk sharing works

The simple framework

How did we do?

☑ Risk sharing works

☒ Ex-Post fairness

The simple framework

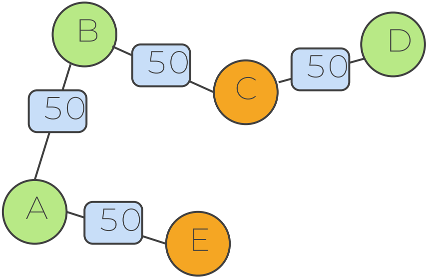
How did we do?

- ☑ Risk sharing works
- ☒ Ex-Post fairness
- ☒ Works for irregular networks

The simple framework

How did we do?

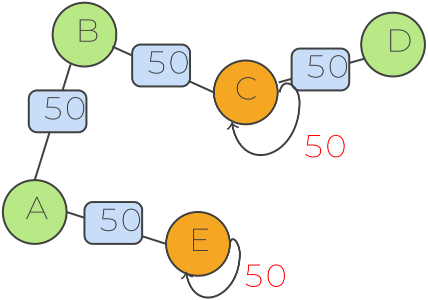
- ☑ Risk sharing works
- ☒ Ex-Post fairness
- ☒ Works for irregular networks



The simple framework

How did we do?

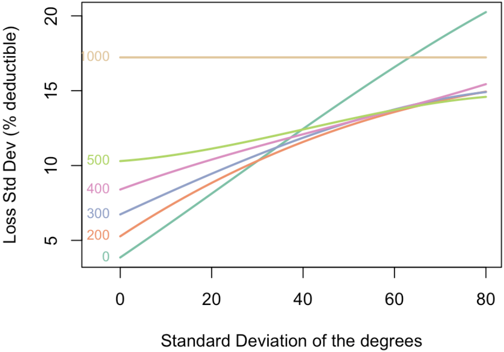
- ☑ Risk sharing works
- ☒ Ex-Post fairness
- ☒ Works for irregular networks



The simple framework

How did we do?

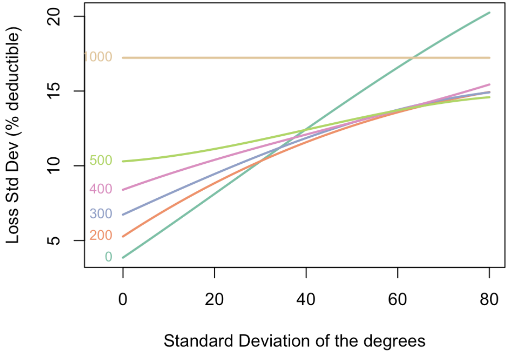
- ☑ Risk sharing works
- ☒ Ex-Post fairness
- ☒ Works for irregular networks



The simple framework

How did we do?

- ☑ Risk sharing works
- ☑ Ex-Post fairness
- ☒ Works for irregular networks



Optimal reciprocal contributions

- Recall: $\xi_i = f(s, Y_i, Z_i, d_i, \gamma)$

Optimal reciprocal contributions

- Recall: $\xi_i = f(s, Y_i, Z_i, d_i, \gamma)$
- Can γ be optimized for any d?

Optimal reciprocal contributions

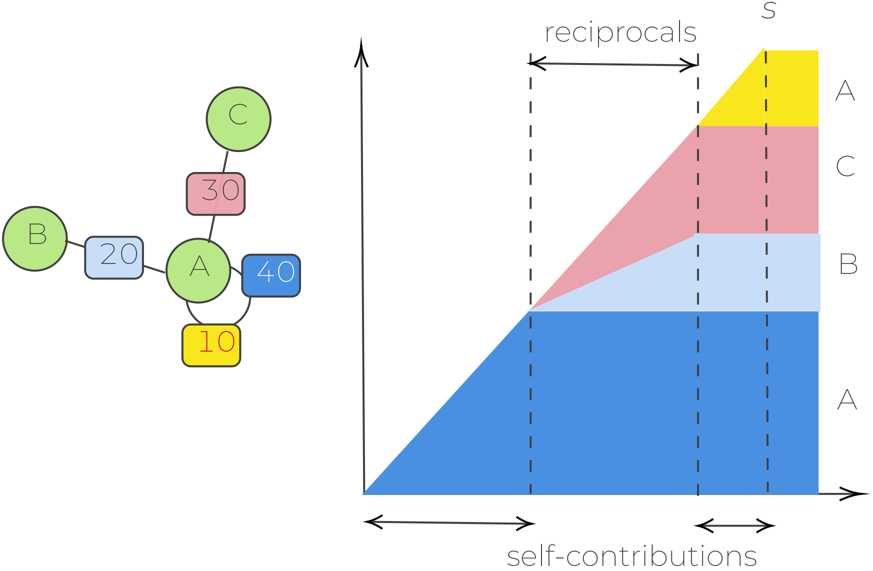
- Recall: $\xi_i = f(s, Y_i, Z_i, d_i, \gamma)$
- Can γ be optimized for any d?
- Here: consider a "global" coverage

$$\boldsymbol{\gamma}_1^* = \operatorname{argmax} \left\{ \sum_{(i,j)\in\mathcal{E}} \gamma_{(i,j)} \right\}$$

$$\text{s.t. } \gamma_{(i,j)} \in [0, \gamma],\ \forall (i,j) \in \mathcal{E}$$

$$\sum_{j\in\mathcal{V}_i} \gamma_{(i,j)} \leq s,\ \forall i \in \mathcal{V}$$

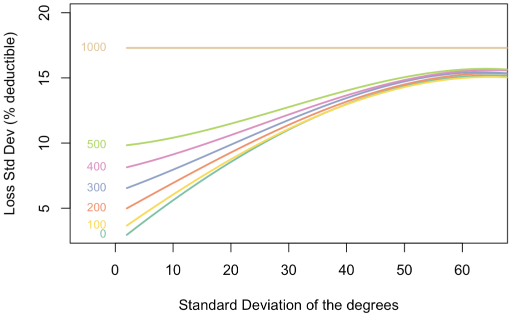
A P2P-Network mechanism



Results

How did we do?

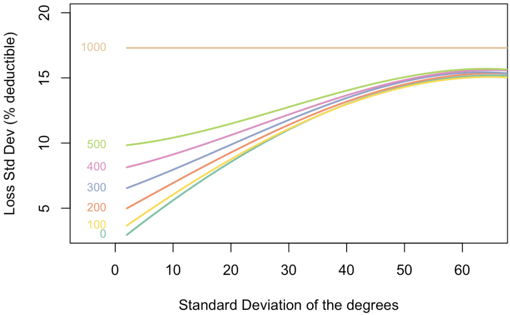
☑ Risk sharing works



Results

How did we do?

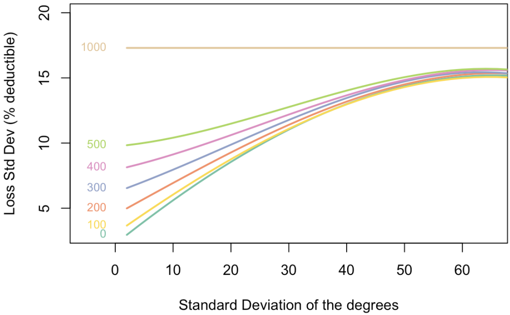
- ☑ Risk sharing works
- ☑ Ex-Post fairness



Results

How did we do?

- ☑ Risk sharing works
- ☑ Ex-Post fairness
- ☑ Works on arbitrary networks



Optimal reciprocal contributions - II

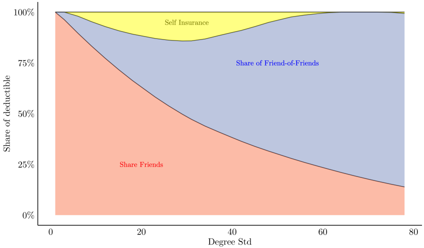
Why stop at friends?
We can also consider friends-of-friends!

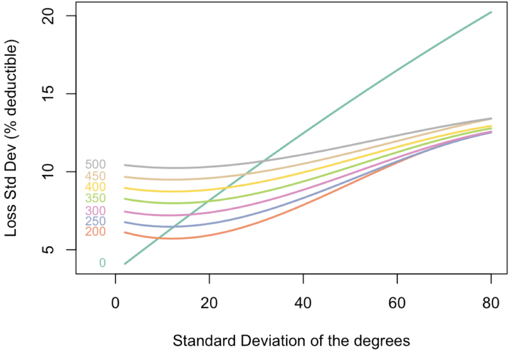
$$\gamma_2^\star = \operatorname{argmax}\left\{\sum_{(i,j)\in\mathcal{E}^{(2)}} \gamma_{(i,j)}\right\}$$

$$\text{s.t. } \gamma_{(i,j)} \in [0,\gamma_2],\ \forall (i,j)\in\mathcal{E}^{(2)},\ \mathcal{E}^{(2)} \text{ from } A^2$$

$$\sum_{j\in\mathcal{V}_i^{(1)}} \gamma^\star_{1:(i,j)} + \sum_{j\in\mathcal{V}_i^{(2)}} \gamma_{(i,j)} \le s,\ \forall i\in\mathcal{V}$$

Optimal reciprocal contributions - II

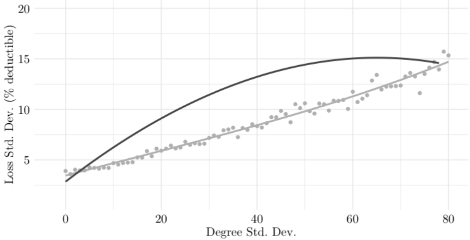




Further Research: Utility based optimization

The mechanism works, but also allows for the incorporation of utility theory (eg. quadratic utility):

$$\min_{w} w^\top Q w$$



This also enables self-insurance or heterogeneous risks! (similar to [Feng et al., 2022], but for arbitrary networks)

Densely connected subgroups I

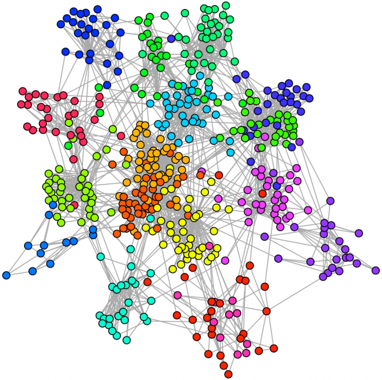
The independence assumption of connections is not realistic - maybe we can partition the graph somehow?[Feng et al., 2022] propose a mechanism also based on cliques, but:

$$\mathbb{V}[\xi] = \mathbb{V}[X]\left(\frac{j}{n}\right)^2$$

with j the number of cliques in a graph, and n the ideally sized clique, making it a *minimal clique problem*, which is NP-complete

Densely connected subgroups II

Instead we can consider clusters (densely connected subgroups)



Similarity between models I

Research on cliques is common in the social sciences eg. [Boucher, 2020] works on networks generated by racial homophily - where the sociodemographic covariates cannot explain friendships fully. But does it work when we have incomplete cliques?

- How do rate classes and CMRS based on homophily relate to each other?
- Can we combine "traditional" GLMs and Network approaches?
- Can we use NN-descent ([Dong et al., 2011]) with the target variable y?

Model Competition in Markets

Motivation I

Introduction of AI has potentially profound impact on markets - with important policy implications - OECD opened up the discussion for AI in insurance and competition [OECD, 2020, OECD, 2021].

Capabilities of ML models get clearer by the year - but effects on markets are still poorly understood.

Source: [OECD, 2020]

Motivation II

Market prices are influenced by availability of new data but also new model types

Effect of machine learning models on a portfolio has been studied [Denuit et al., 2021, Wüthrich, 2022]

But usually effects in markets is scarcely studied, as they are notoriously complicated and opaque..

Munich RE

Digitalisation

How Automated Machine Learning is reshaping insurance pricing

Artificial intelligence is changing our lives in many ways. Insurers can use data and algorithms to predict the behaviour of their customers. This also offers new opportunities for underwriting and pricing. Massimo Cavadini, Head of Actuarial Consulting & Data Analytics at Munich Re, explains how automated machine learning (AutoML) puts the pricing expert at the heart of the decision-making process and how clients benefit from this.

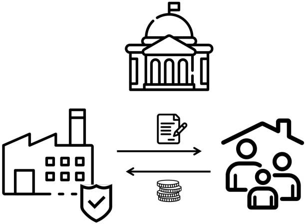
14.02.2022

Source: [Munich Re, 2022]

Motivation III

...But we can run a digital field experiment! Together with the CPG (@Imperial College) this was done in 2021.

The idea builds upon [Charpentier et al., 2015] but with more data, better control and more participants. This allows us to **study the impact of new models on consumers, producers and the regulator**.

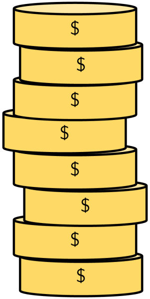


Literature I

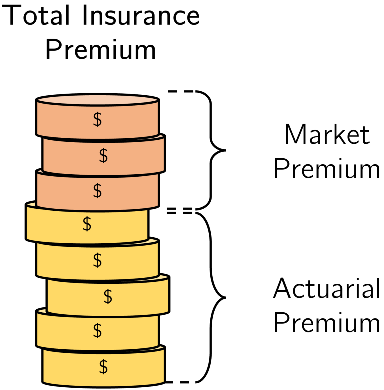
For models themselves, there is contradicting evidence to their performance [Noll et al., 2020, Dugas et al., 2003], especially in insurance. But in theory the performance *capabilities* are quite clear see eg. [Hastie et al., 2009].

In theory, higher pricing precision should mean lower consumer prices, but case studies found *higher* supplier margins [Assad et al., 2020, Vogell et al., 2022]. In general, if there is an analysis that goes beyond a theoretical model, it is usually a single-case studies.

Total Insurance Premium

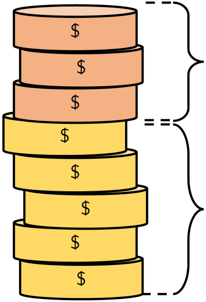


Premium Decomposition & Research Questions



Premium Decomposition & Research Questions

Total Insurance Premium



Market Premium

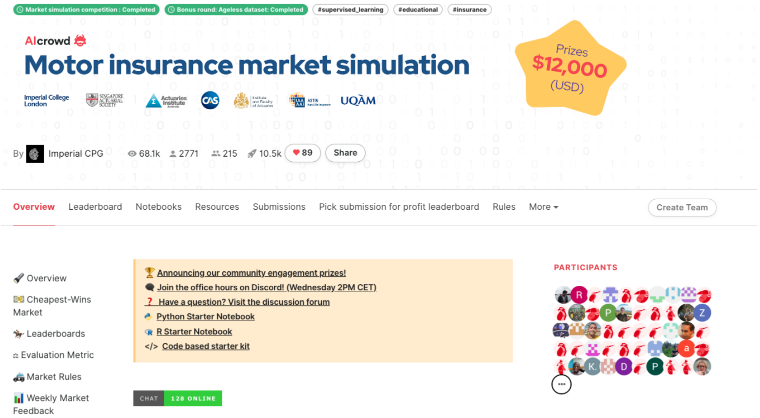
Actuarial Premium

Effect of Pricing-Strategy?

Complex models, better predictions?

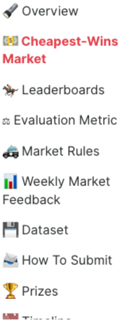
How are different segments affected?

Setup I



aicrowd.com/challenges/insurance-pricing-game

Setup II



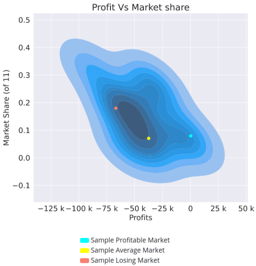
	Company 1	Company 2	Claim amount
Policy 1	120	90	100
Policy 2	30	25	10
Policy 3	10	15	0
Policy 4	5	10	0
Total revenue	15	115	
Total loss	0	110	
Total profit	15	5	

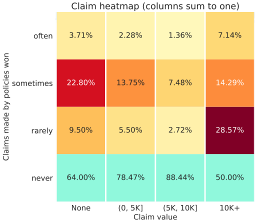
Most profit

aicrowd.com/challenges/insurance-pricing-game

Setup III

The competition goes on for 11 weeks, and importantly, there is feedback:





Driver Summaries

Policies won	drv_sex1	drv_age1	drv_age_lic1	drv_drv2	drv_sex2	drv_age2	drv_age_lic2
often: 54.5 - 100.0%	M (63%)	61.57	38.70	No (73%)	F (66%)	50.00	27.74
sometimes: 18.2 - 54.5%	M (65%)	59.19	36.87	No (72%)	F (63%)	51.05	29.14
rarely: 9.1 - 18.2%	M (63%)	60.92	38.35	No (71%)	F (62%)	51.09	29.12
never:	M (58%)	58.17	36.08	No (64%)	F (60%)	51.16	29.31
in sample profitable market	M (64%)	61.05	38.33	No (74%)	F (64%)	49.29	27.25
in sample average market	M (65%)	61.73	38.99	No (75%)	F (65%)	50.43	28.29
in sample losing market	M (63%)	59.89	37.37	No (70%)	F (64%)	50.70	28.69

Setup IV

Participants compete in against nine other contenders for customers looking for a third party liability insurance (think: a consumer that uses a comparison site)
All possible $\binom{n}{10}$ markets are considered and we can simulate outcomes for visualisations.

Source: [LowestRates.ca, 2022]

Research Plan I

- ☑ Get overview of data, analyse statics, reproduce results
- ☐ Understand and quantify model complexity (in terms of output complexity)
- ☐ Understand market dynamics
  - ☐ In simulated environment
  - ☐ In competition environment

Then go back to the three research areas

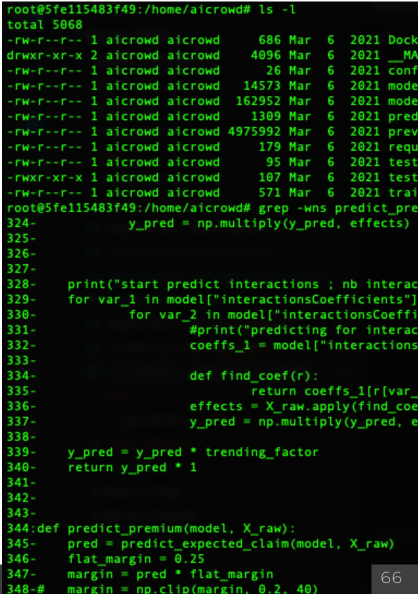
Research Plan II

Other projects that can be accomplished on the data:

- ☐ Explore combination of models & incomplete information
  - ☐ In a static world
  - ☐ In a dynamic world ([Zheng et al., 2020])

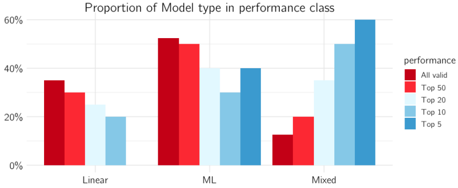
State of play I

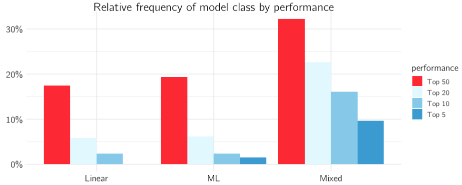
Over 11 weeks, ≈ 2,000 participants submitted a total of 10,000 models Over 80% of participants devoted more than two hours weekly to modelling in the first phase. Since we have the model code, we can classify them into Linear, Machine learning or Mixed models.

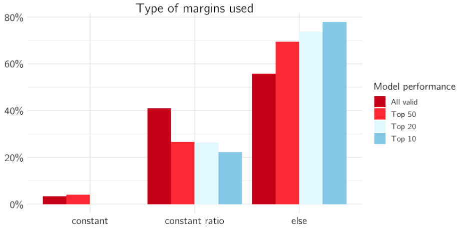


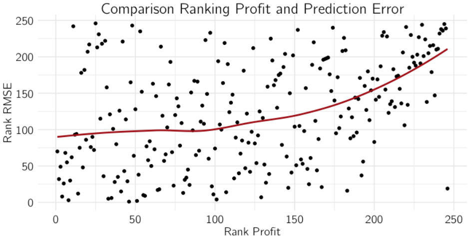
State of play II

Here: 250 valid submissions from the final week on test-data

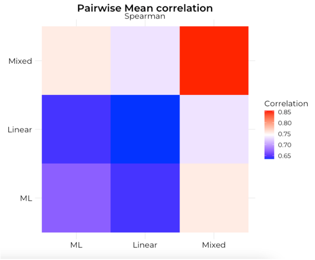


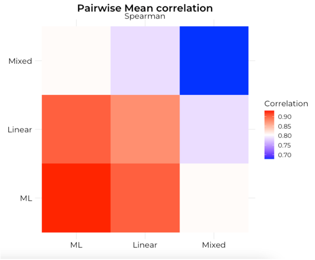


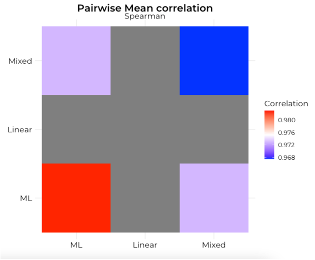




Some insights I

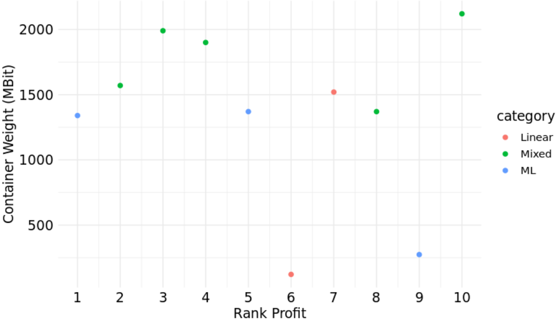






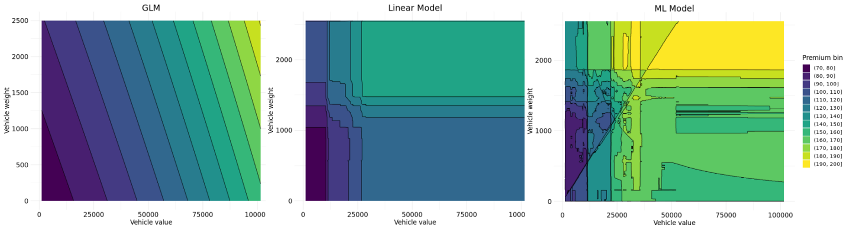
Some insights II

Is (model) complexity really what we are after?



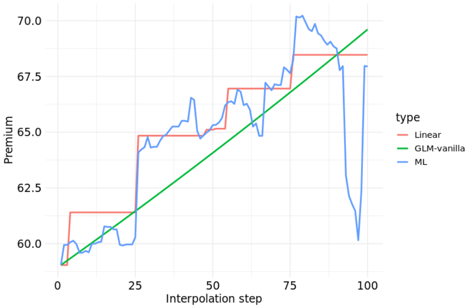
There is currently few existing research on this, especially for the regression case..

Some insights III



$$
\begin{aligned}
\boldsymbol{x}_i &= (x_{i,1}, \ldots, x_{i,D}), \quad \forall i \in \mathcal{D} \\
\boldsymbol{\mu} &= (\mu_1, \ldots, \mu_D) \\
\boldsymbol{v}_i &= (\mu_1, \ldots, \mu_{j-1}, u, \ldots, \mu_{l-1}, t, \ldots, \mu_D) \\
u &\in [\min\{x_{\cdot,j}\}, \max\{x_{\cdot,j}\}], t \in [\min\{x_{\cdot,l}\}, \max\{x_{\cdot,l}\}]
\end{aligned}
$$

Some insights IV



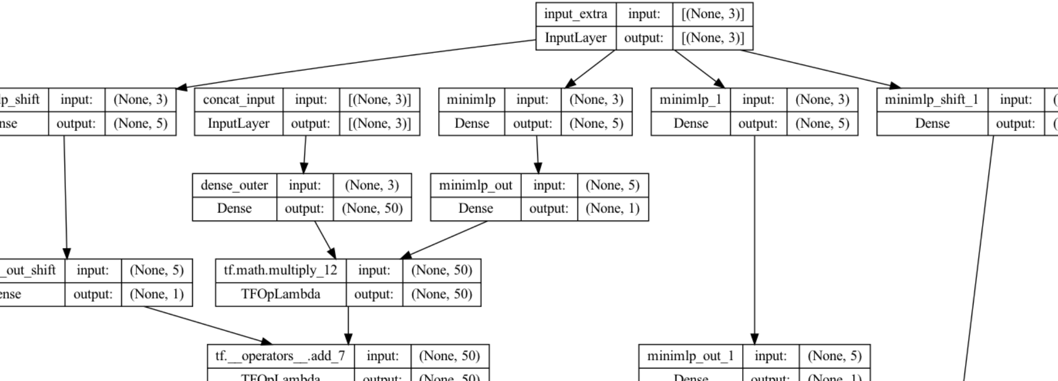
$$\boldsymbol{v}_i = (1-\lambda_i)\boldsymbol{x}_1 + \lambda_i \boldsymbol{x}_2, \quad \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{D}, \lambda_i \in [0,1]$$

$$\hat{y}_{\mathcal{M}} = f_{\mathcal{M}}(\boldsymbol{v}_i)$$

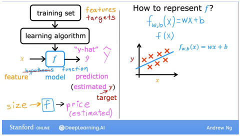
Next Steps

- ☐ Model Quantification
  - ☐ Expand search beyond a few models
  - ☐ Expand into 2d space around points & quantify surface
- ☐ Understand market dynamics (in progress)

Statistics of Deep Learning



Motivation



Source: [Coursera, 2022]

Source: [The Economist, 2016]

There is probably some middle ground between linear regression and the fully automatic *Maschinenmensch*. Evidence points to combination of "AI" and "classic" statistics as most successful pairing.

Literature I

Often cited are [Hornik et al., 1989, Leshno et al., 1993], but beyond really specific tasks such as speech and images, deep learning remains a niche product. In general, there is rather mixed evidence from the field and usage of ML-only approaches ie. [Makridakis et al., 2020]. Research and competitions also point in the direction of hybrid approaches as the best performers ie. M4, M5, NFL Big Data Bowl. However, a good mathematical understanding of the problem is key here.

We can look to theoretical approaches for guidance, for example [Bühlmann and Yu, 2003] and stop considering ML as something separate to the "usual" statistics.

Outline

In general, AI is said to transform the world with new models and new data. If we understand how ML algorithms work, we can start to integrate them into more standard models but we can also combine different sources of data.

Insights based on work for *Deep Kaplan-Meier* (2022) - under review, *Nonparametric sieve estimation* [Ratz, 2022] - presented at VieCo (2022) and at CFE/CMS (in two weeks) and *Multitask estimation* - working paper.

The approximation view I

Usually, model observations as a random sample from a probability distribution which can be expressed in terms of parameter θ, such that the distribution falls within a family $\{f(\cdot;\theta)|\theta \in \Theta\}$, usually Θ is finite dimensional and the functional form for f is restricted.

$$y = X\beta + \varepsilon, \quad \Theta = \{(\beta,\sigma) : (\beta,\sigma) \in \mathbb{R}^D \times \mathbb{R}^+\}$$

Quite flexible as for real valued, continuous f on $[a,b]$, $\forall\varepsilon > 0$ there exists a polynomial p such that $\forall x \in [a,b], |f(x) - p(x)| < \varepsilon$ (Weierstrass approximation theorem).

The approximation view II

Usually not really efficient though..
Instead consider an estimator of the form:

$$f(x,\theta) = \beta_0 + \sum_{k=1}^{m} \beta_k G(x^\top \gamma_k), \tag{1}$$

for some function G and $\theta = (\beta_0, \ldots, \beta_m, \lambda_1, \ldots, \lambda_m)$. Then for a suitably chosen G, [Hornik et al., 1989] and [Leshno et al., 1993] showed that (1) can approximate any function (for example in $C(\mathbb{R}^D, \mathbb{R})$), arbitrarily well if m is chosen large enough.

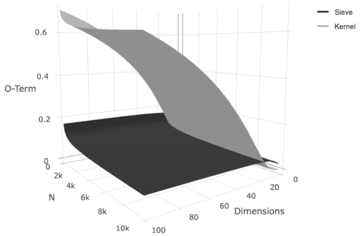
The approximation view III

The issue becomes that we need to optimize some criterion to find the parameters. Now if Θ is an infinite dimensional parameter space, maximizing an optimization criterion is usually no longer well defined. [Grenander, 1981] proposed instead to estimate the problem using a sequence of approximation spaces Θ_n which are dense in Θ. Usually $(\Theta_n \subseteq \Theta_{n+1} \dots)$ such that there is some π_n and metric d, such that $\forall \theta \in \Theta, \exists \pi_n,$ s.t. $d(\theta, \pi_n \theta) \to 0$ as $n \to \infty$ and π_n a projection mapping from Θ to Θ_n (see eg. [Chen et al., 2001])

TL;DR We can pretend to approximate our true function in a simpler approximation space. This then allows us to treat neural networks as nonparametric estimators.

Rates of convergence

Given that this is an approximation problem, we can derive rates of convergence. This was done for neural networks with "squashing" functions in [Chen et al., 2001] and extended to ReLU functions in [Ratz, 2022]. Usually, this means balancing bias and variance of the approximation and allows us to draw conclusions from the rates:



A flexible Kaplan-Meier Estimator I

We can add flexibility to many models by changing the form slightly. In survival estimation, some work has already been done. For example, [Yu et al., 2011] changed the Cox-PH model to allow a flexible hazard function. [Fotso, 2018] then extended this to include neural networks.

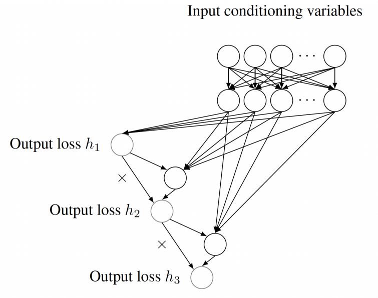
Here we consider a slightly different parametrization, for K discrete time bins:

$$Y_{i,j} = \begin{cases} 1 \text{ if } \tilde{\tau}_i < j \\ 0 \text{ otherwise} \end{cases} \qquad \forall i, j = 0, 1, \ldots, K$$

$$W_{i,j} = \begin{cases} 1 \text{ if } c_i < j \\ 0 \text{ otherwise} \end{cases} \qquad \forall i, j = 0, 1, \ldots, K$$

A flexible Kaplan-Meier Estimator II

Now instead, assume we try to predict survival up to a point *j* with a binary regression. But we link their results in a neural architecture



A flexible Kaplan-Meier Estimator III

Then we can set up the Kaplan Meier estimator as the solution to the following weighted likelihood function:

$$\hat{z}_j = \arg\max_{z_j} \prod_{i=1}^{n} z_j^{w_{i,j} y_{i,j}} (1 - z_j)^{w_{i,j}(1 - y_{i,j})} \tag{2}$$

we can then model z_j with an arbitrary function dependent on covariates and approximate it with a neural network! This leads to a few interesting propositions:

A flexible Kaplan-Meier Estimator IV

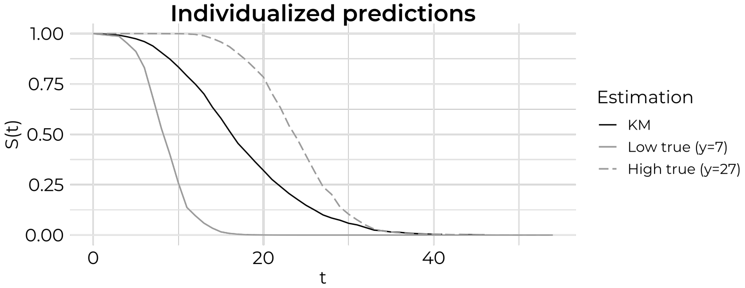
Equivalence to Kaplan-Meier

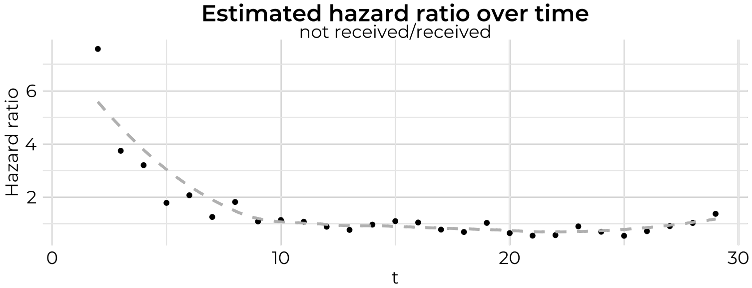
The Kaplan-Meier estimator is the expected value of the n individual predictions from the model specified by equation (2).

Consistency under informative censoring

If conditional independence between censoring and event time holds, the CKM estimator converges to the true hazard rate, even if censoring and event time are dependent.

A flexible Kaplan-Meier Estimator V




Estimated hazard ratio over time
not received/received
Hazard ratio
6
4
2
0
10
20
30
t

Task Combination I

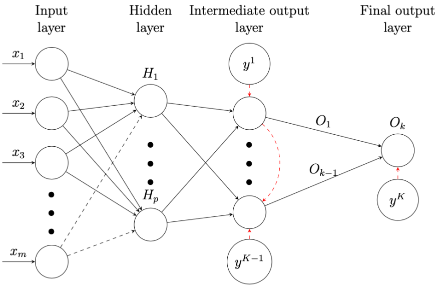
What if we push the universal approximation theorem? It remains valid even for multiple tasks (if we accept some restrictions on discontinuity)

[Dosovitskiy and Djolonga, 2019] explore multi-task learning when there is an inherent trade-off between the tasks. But what if we instead combine the tasks? - Some preliminary evidence

Task Combination II

An example (premium prediction)
$A|\boldsymbol{X} \sim \mathcal{B}(p_{\boldsymbol{X}})$ and $S|\boldsymbol{X} \sim \mathcal{G}(\mu_{\boldsymbol{X}})$

$$\hat{\pi}(\boldsymbol{x}) = \hat{\mathbb{E}}[A|\boldsymbol{X} = \boldsymbol{x}] \times \hat{\mathbb{E}}[S|\boldsymbol{X} = \boldsymbol{x}]$$



So we are mostly interested in the combination, but divide the tasks as they are more manageable. Can we use this somehow?

Task Combination III

Let $\mathcal{L}^k(\cdot,\cdot)$ be the loss function associated to task k. For n observations and task weights λ_k,

$$\hat{\theta}^* = \arg\min_{\theta} \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k \mathcal{L}^k(y_{i,k}, f(x_i, \theta, \lambda_k)),$$

where f is approximated with a FFNN

Task Combination IV

Formally, the basis for loss conditioning is based on YOTO [Dosovitskiy and Djolonga, 2019]

$$\arg\min_{\theta} \mathbb{E}_{x,y} \mathcal{L}(y, F(x, \theta), \lambda) \tag{3}$$

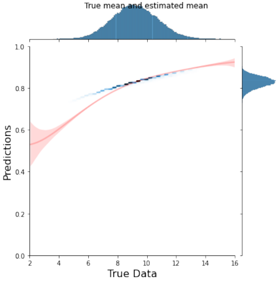
Here the model is optimized independent from λ, assume it is fixed and a hyperparameter to tune

$$\arg\min_{\theta} \mathbb{E}_{x,y} \mathbb{E}_{\lambda} \mathcal{L}(y, G(x, \theta, \lambda), \lambda) \tag{4}$$

This time we make the model "learn" to optimize over a distribution of λ which then allows to condition on it

Task Combination V

In theory, we can recover an arbitrary relationship between the sub-parts of the model, and it seems to work in practice.



Next Steps

- ☐ Approximation Results
    - ☐ Derive approximation using the theory of sieves
    - ☐ Derive Rademacher Inequality
- ☐ Rerun graphs and submit :)

Appendix

Conditional Mean Risk sharing details I

Although established as optimal (in the convex order sense), the conditional mean risk sharing of [Denuit and Dhaene, 2012] has an important drawback: it is only guaranteed if $X_i|S$ is known (that is, the joint distribution of $\boldsymbol{X} = (X_1, \ldots, X_n)$ is known.

In eg. [Denuit and Robert, 2020b] this is extended and show preference relations over different risk pools. Further, they show that increasing participants in a pool is always beneficial under conditional mean risk sharing.

[Denuit and Robert, 2020a] is then concerned with modelling the conditional dependence structure, which are represented as a graphical model.

Optimization approach of Feng I

[Feng et al., 2022] then take a similar approach, supposing that $\mathbb{V}[h_i(S)] \leq \mathbb{V}[X_i], \forall i$ is the goal of the sharing mechanism. The novelty here is to consider:

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i, \quad Y_i = h_i(X_1, \ldots, X_n)$$

with the special case $h_i(X_1, \ldots, X_n) = h_i(S)$. Their results are devised for quota sharing (which is similar to what we consider in our paper as well), that is for weights $\alpha_{i,j}, \sum_{j=1}^{n} \alpha_{i,j} = 1$

Optimization approach of Feng II

[Abdikerimova and Feng, 2021] consider P2P transfer networks. If p_i is the probability that i suffers a loss, one should impose:

$$\alpha_{i,1} p_1 + \cdots + \alpha_{i,n} p_n = p_i$$

Their main results are then centered around specific solutions for different circumstances. Note that this is also similar to our optimization problem.

The Balance property I

In the words of [Wüthrich, 2022], even if individual error is small, the aggregated portfolio level error might be substantial. With the example of a classification problem:

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \sigma(\hat{\beta}_0^{\mathrm{MLE}} + x_i^{\mathsf{T}}\hat{\beta}^{\mathrm{MLE}}) = \frac{1}{n}\sum_{i=1}^{n} \hat{p}^{\mathrm{MLE}}(x_i)$$

taking expectations then gives the unbiasdness property on the population level. This holds for GLMs within the exponential family with canonical links. Note though, that this does not hold for not-fully trained ANNs.

Appendix

Image on section title for Risk sharing on Networks from: [Föhn-Gasser, 1952], Icons not specifically referenced are from flaticon.com

References I

Abdikerimova, S. and Feng, R. (2021).
Peer-to-peer multi-risk insurance and mutual aid.
European Journal of Operational Research.

Assad, S., Clark, R., Ershov, D., and Xu, L. (2020).
Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market.

Boucher, V. (2020).
Equilibrium homophily in networks.
European Economic Review, 123:103370.

Bühlmann, P. and Yu, B. (2003).
Boosting with the l 2 loss: regression and classification.
Journal of the American Statistical Association, 98(462):324–339.

References II

Charpentier, A., Denuit, M., Elie, R., et al. (2015).
Segmentation et mutualisation, les deux faces dune même pièce?
Risques, 103:57–64.

Charpentier, A., Kouakou, L., Löwe, M., Ratz, P., and Vermet, F. (2021).
Collaborative insurance sustainability and network structure.
arXiv preprint arXiv:2107.02764.

Chen, X., Racine, J., and Swanson, N. R. (2001).
Semiparametric arx neural-network models with an application to forecasting inflation.
IEEE Transactions on neural networks, 12(4):674–683.

Coursera (2022).
Machine learning, andrew ng.
coursera.com.

References III

Denuit, M., Charpentier, A., and Trufin, J. (2021).
Autocalibration and tweedie-dominance for insurance pricing with machine learning.
Insurance: Mathematics and Economics, 101:485–497.

Denuit, M. and Dhaene, J. (2012).
Convex order and comonotonic conditional mean risk sharing.
Insurance: Mathematics and Economics, 51(2):265–270.

Denuit, M. and Robert, C. (2020a).
Conditional mean risk sharing for dependent risks using graphical models.
LIDAM Discussion Papers ISBA, (2020029).

Denuit, M. and Robert, C. (2020b).
Risk reduction by conditional mean risk sharing with application to collaborative insurance.
Technical report, Université catholique de Louvain, Institute of Statistics, Biostatistics and .

References IV

Dong, W., Moses, C., and Li, K. (2011).
Efficient k-nearest neighbor graph construction for generic similarity measures.
In *Proceedings of the 20th international conference on World wide web*, pages 577–586.

Dosovitskiy, A. and Djolonga, J. (2019).
You only train once: Loss-conditional training of deep networks.
In *International conference on learning representations*.

Dugas, C., Bengio, Y., Chapados, N., Vincent, P., Denoncourt, G., and Fournier, C. (2003).
Statistical learning algorithms applied to automobile insurance ratemaking.
In *CAS Forum*, volume 1, pages 179–214. Citeseer.

Feng, R., Liu, C., and Taylor, S. (2022).
Peer-to-peer risk sharing with an application to flood risk pooling.
Annals of Operations Research, pages 1–30.

References V

Forbes (2018).
The Top 5 Industries Most Hated By Customers.
Forbes, Blacke Morgan.

Fotso, S. (2018).
Deep neural networks for survival analysis based on a multi-task framework.
arXiv preprint arXiv:1801.05512.

Föhn-Gasser, X. (1952).
Hinterer heubrig, muotathal: Wildheuer beim zirknen.
[Online; accessed Nov 25, 2022].

Granovetter, M. (2005).
The impact of social structure on economic outcomes.
Journal of Economic Perspectives, 19(1):33–50.

References VI

Grenander, U. (1981).
Abstract inference.
Technical report.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009).
The elements of statistical learning: data mining, inference, and prediction, volume 2.
Springer.

Hornik, K., Stinchcombe, M., and White, H. (1989).
Multilayer feedforward networks are universal approximators.
Neural networks, 2(5):359–366.

Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S. (1993).
Multilayer feedforward networks with a nonpolynomial activation function can approximate any function.
Neural networks, 6(6):861–867.

References VII

Lin, M., Prabhala, N. R., and Viswanathan, S. (2013).
Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending.
Management science, 59(1):17–35.

Liu, Z., Shang, J., Wu, S.-y., and Chen, P.-y. (2020).
Social collateral, soft information and online peer-to-peer lending: A theoretical model.
European Journal of Operational Research, 281(2):428–438.

LowestRates.ca (2022).
Premium calculator.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2020).
The m4 competition: 100,000 time series and 61 forecasting methods.
International Journal of Forecasting, 36(1):54–74.

References VIII

Munich Re (2022).
How automated machine learning is reshaping insurance pricing.
munichre.com.

Noll, A., Salzmann, R., and Wuthrich, M. V. (2020).
Case study: French motor third-party liability claims.
Available at SSRN 3164764.

OECD (2020).
The impact of big data and artificial intelligence (AI) in the insurance sector.
OECD Paris, France.

OECD (2021).
AI in Business and Finance.
OECD Paris, France.

Ratz, P. (2022).
Nonparametric value-at-risk via sieve estimation.
arXiv preprint arXiv:2205.07101.

References IX

The Economist (2016).
March of the machines.
economist.com.

Vogell, H., Coryne, H., and Little, R. (2022).
How a secret rent algorithm pushes rents higher.

Wüthrich, M. V. (2022).
The balance property in neural network modelling.
Statistical Theory and Related Fields, 6(1):1–9.

Yu, C.-N., Greiner, R., Lin, H.-C., and Baracos, V. (2011).
Learning patient-specific cancer survival distributions as a sequence of dependent regressors.
Advances in neural information processing systems, 24.

References X

Zheng, S., Trott, A., Srinivasa, S., Naik, N., Gruesbeck, M., Parkes, D. C., and Socher, R. (2020).
The ai economist: Improving equality and productivity with ai-driven tax policies.
arXiv preprint arXiv:2004.13332.