Data Paucity in New Insurance Markets

Anirudha Balasubramanian

March 2, 2022

Abstract

Why are insurance markets for new or nonstationary risks often tiny, even when those risks are massive? To answer this question, I develop a general formalism ("CUE") for dynamic non-Bayesian learning and decisionmaking where decisionmakers penalize model uncertainty. First, I show that data paucity may stop a CUE insurer from ever underwriting, thus creating a "data trap." Second, I narrow to a specific model where insurers use nonparametric $(1 - \alpha)$-confidence intervals to assess prospective policies and show the consequent intrinsic margin reduction of insurance. In particular, I show how policy limits and deductibles emerge as features of equilibrium contracts even without adverse selection or moral hazard. Finally, I discuss implications for datasharing and antitrust policy in young insurance industries.

*Stanford Graduate School of Business. Email: balasub@stanford.edu. I thank Levent Alpöge, Michael Celentano, Daniel Chen, Benjamin Hébert, Yucheng Liang, Suraj Malladi, Mike Ostrovsky, Peter Reiss, Amit Seru, Andrzej Skrzypacz, Nilesh Tripuraneni, and David Yang for useful discussions. All errors are my own.
1 Introduction

“[C]ompared with other risks covered by insurance...cyber risk is perhaps the fastest-evolving and least understood. A big challenge faced by the market is the scarcity of data on past incidents. The importance of modeling cannot be understated when it relates to pricing risk, and the lack of historical data and unpredictability of risk make it difficult to model and therefore underwrite. There is a significant qualitative aspect to pricing that insurers rely on when pricing policies...The adverse effect of insufficient data is that insurance firms struggle to price cyber risk...the cyber insurance market is relatively small in size [due] to the relatively high premiums for relatively limited coverage.”

“We can figure the probability of a quake or a hurricane but don’t know as much in cyber. It’s uncharted territory on the insurance side and will get worse, not better.”

At an estimated $600 billion to $1 trillion of cost annually (McAfee-CSIS, 2018), global cyber risks loom large. Corporations rank it as the top peril they face (Allianz, 2022). But the insurance market for these risks is extremely small, at under $10 billion of written premiums globally; insofar as policies are written, loss limits are low and prices – relative to incurred losses – are high (NAIC, 2021). Although cyber may be the starkest example, undersized insurance and lending markets for new, nonstationary, or otherwise data-scarce risks are common.¹

In this paper, I argue that data paucity – the lack of appropriate data with which to estimate the risk distribution – can be an important barrier to insurability in such settings. This claim is hard to square with a standard Bayesian model of insurance and a risk neutral insurer²: lack of data – manifesting as higher uncertainty in priors – increases the gains from trade and both intrinsic and extrinsic margins of coverage for a fixed level of adverse selection or moral hazard. Bayesian decisionmakers simply collapse uncertainty into variance, which these models assume is diversifiable.

¹Climate change has increased the frequency and severity of secondary perils like winter storms, wildfires, windstorms, and severe flooding. Most associated losses are uninsured (Rauch and Allman, 2020). In certain geographies, recent years have seen reductions in policy limits and coverage in associated markets, with uncertainty, realized losses, and occasional price controls jointly responsible.

²This simplification is often made as reduced form, for an insurer with the ability to diversify risks.
How does data paucity hamper insurability? Suppose insurers penalize (model) uncertainty, treating it differently from known risks. Indeed, such uncertainty is nondiversifiable across claims. Lack of data results in higher model uncertainty about the underlying distribution of claims or aggregate portfolio payout $F$. This uncertainty imposes a cost on the insurer and may drive the insurer out of the market entirely. The insurer’s participation decision has dynamic implications for the market if underwriting is the only reliable non-governmental way to collect loss data necessary for training models: the market may stay “stuck” in a data trap.\(^3\) Even if the insurer chooses to underwrite, she may distort the intrinsic margin of contracting activity, as the contract the she chooses shapes the uncertainty she faces.

Specifically, in Section 2, I develop a general framework for “conservative uncertainty evaluation” for a dynamic decisionmaker with recursive preferences who makes decisions based on statistical information (dataset of iid draws) available to him at time $t$. Conservatism is defined relative to a benchmark risk neutral decisionmaker with a possibly misspecified Bayesian model for the world. As such, my model imposes no notion of “belief” or “prior” on the decisionmaker and accommodates both purely algorithmic decisionmaking as well as more conventional theories of ambiguity aversion. In the former case, starter datasets may nonparametrically determine the uncertainty set that economic theory typically takes as given. I cast such a decisionmaker as a monopolist insurer making take-it-or-leave-it offers to a mean-variance utility client with the same Bayesian model as the benchmark decisionmaker. In Theorem 2.1, I show how insurance may never occur in the presence of a finite starter dataset $D$ and arbitrarily small fixed costs, and in Theorem 2.2 show how expanding the dataset sufficiently will, with high probability, cause the market to function.

In Section 3, I define a nonparametric notion of conservative uncertainty evaluation and introduce the conservative confidence bound (CCB) model as an example. A CCB insurer uses a nonparametric $(1 - \alpha)$ Dvoretzky-Kiefer-Wolfowitz (DKW) envelope around the CDF (Massart, 1990) constructed from the dataset to construct confidence bounds for the expected loss from any given contract: the upper bound of such a loss is the insurer’s perceived cost basis. Regardless of the buyer’s bounded valuation, it is obvious that the insurer must offer a contract with a loss limit. When the buyer’s valuation is empirical mean-variance, based also on the dataset, I show that the market can fail even without fixed costs. In Theorem 3.1 I characterize the optimal one-period contract as a function of the dataset: it contains a loss

\(^3\)Friedman and Thomas (2017) of Deloitte write of a “vicious circle” in cyberinsurance: lack of data leads to a lack of underwriting, which in turn precludes the data collection necessary for future underwriting.
limit, may contain a deductible, and features no coinsurance. This exemplifies the intrinsic margin reductions of insurance that CUE can create and matches qualitative features of cyberinsurance contracts. In Section 3.3, I argue that the CCB model does so better than moral hazard or adverse selection models and discuss how, if supplemented with slow moving capital, it tenders realistic empirical predictions about the evolution of policy limits over time.

In Section 4, I study both government investment in data and mandated datasharing as prospective policy measures. While there is no formal inefficiency in the CCB model, I show in Proposition 4.1 that the government – even facing higher costs – may wish to acquire data when an insurer would not. Specifically, because the value of a prospective data investment is uncertain, such a situation can arise if the government, in its purchase decision, penalizes this uncertainty differently from how the insurer would. In Section 4.2, I develop a reduced form model of two Bertrand-competing insurers, who use data as a cost-reducing input in two similar markets. I show how they may create social surplus by voluntarily sharing anonymized or binned loss data and reducing their respective uncertainty. Proposition 4.2 demonstrates that they share too little with each other, meaning that a government mandate – or relaxation of antitrust rules – can create efficiency. While an intuitive and colloquially understood point, absent the obvious dynamic concerns, uncertainty reduction typically takes a back seat to inferential concerns and adverse selection in economic analysis of (non-anonymized) datasharing.

1.1 Related Literature

CUE, as defined in Section 2, relates closely to a sizeable literature on dynamic decision-making under uncertainty. My formulation of the data generating process is subsumed by Epstein and Schneider (2003, 2007). CUE itself nests smooth ambiguity (Klibanoff et al., 2005, 2009) and maxmin expected utility (Gilboa and Schmeidler, 1989) Perhaps most similar to the CCB model are the quantile maximization models of Rostek (2010) and de Castro and Galvao (2019, 2018), while the IDP model relates to the work on unawareness and ambiguity in Grant et al. (2019). To these papers I add a non-Bayesian and nonparametric perspective. First, I allow a flexible valuation $V$ (which only in the limit converges to conditional expectation) and define a purely algorithmic notion of CUE with recourse to a limit of Bayesian models. Second, the CCB model I develop is a data-driven, closed-form nonparametric specification for uncertainty sets; this new approach circumvents ad-hoc selection of parametric models and priors and is easy to interpret.
Section 3’s results on the CCB model add to an agenda characterizing optimal insurance contracts and sources of market failure. Taken jointly, Raviv (1979), Young and Browne (1997), and Fluet and Pannequin (1997) rationalize such contract features as deductibles, coinsurance, and limits (for low type agents) as responses to adverse selection. Chade and Schlee (2012) offers predictions about the equilibrium risk premium charged to each risk type by a monopolist. Shavell (1979) and Winter (1992) show how moral hazard, similarly, may cause deductibles and coinsurance. A distinct line of work considers ambiguity aversion in insurance contracting. Gollier (2014) solves the optimal insurance contract when a buyer has smooth ambiguity aversion. My paper instead considers uncertainty averse insurers, in line with the survey results of Kunreuther et al. (1993); Theorem 2.1 makes mathematically precise their speculation that ambiguity attitudes may cause market failure. Dietz and Walker (2019) studies the interaction between ambiguity attitudes and capital constraints for insurers, rationalizing why uncertain risks carry higher premiums. In this paper, I show the contractual implications of insurer uncertainty and contribute the argument that data paucity creates a form of nondiversifiable model risk.

Finally, a much newer literature considers the connections between insurance and new data sources, along with the prospect of data sharing. Liang and Madsen (2020) considers “data linkages” used to connect agent behavior across different domains and studies their effect on the useful effort agents exert. Jin and Vasserman (2019) considers data sharing among auto insurers, though it is focused entirely on the equilibrium effects (through an empirical model of moral hazard and adverse selection) of better monitoring technology. He et al. (2020) and Padilla and Pagano (1997, 2000) both study datasharing in lending markets, with the former analyzing lending market competition when borrowers can choose to share their data, and the latter rationalizing the existence of credit bureaus. Lieberman et al. (2019) computes the distributional consequences of a mass data deletion event in a Chilean credit bureau using an adverse selection model. My paper is the only one that studies anonymized data, and I provide a new reason – intimately connected to young markets – for sharing such data: uncertainty reduction.

2 General Model

In this section I define what it means for a decisionmaker to have “conservative uncertainty evaluation” (CUE) relative to a Bayesian benchmark for a class of problems, discuss the applicability of CUE to insurers, and apply this to a monopolist insurer’s underwriting problem to show how CUE can lead to market failure.
2.1 Conservative Uncertainty Evaluation

In Section 2.1.1, I define conservative uncertainty evaluation with respect to a static decision problem. In Section 2.1.2, I extend this definition to specific intertemporal settings.

2.1.1 Static CUE

Let $\mathcal{X} \subset \mathbb{R}$ be some interval of values. Let some unknown $F$ be a distribution function supported on $\mathcal{X}$, with an infinite sequence of “observations” $X_1, \ldots, X_n, \ldots \sim F$. Finally, denote by $D = \{X_1, \ldots, X_n\}$ a dataset of hitherto known/revealed observations. Call these physical primitives.

Define a **Bayesian model** $(\pi, \Theta)$ for the data generating process to consist of (i) $\Theta$, a metric space indexing possible distribution functions, (ii) $\{F^\theta(x)\}$ a set of distribution functions supported on (subsets of) $\mathcal{X}$ such that $F \in \{F^\theta(x)\}$, (iii) $\pi$ a probability measure on $\Theta$ that assigns nonzero mass to any ball surrounding the corresponding “true” $\theta$, (iv) Nature’s draw of a “true parameter” $\theta^N \sim \pi$, and (v) Nature’s subsequent draw of “observations” $X_1, \ldots, X_n, \ldots \sim F^{\theta^N}$.

In a **static decision problem** a DM, given some dataset of iid observations $D = \{X_1, \ldots, X_n\}$ must cardinally evaluate a known deterministic function $g : \mathcal{X} \to \mathbb{R}$ of some unknown independent draw $X \sim F^{\theta^N}$. That is, a decision maker must specify a “certainty equivalent” to some stochastic payoff $g$ given the data $X_1, \ldots, X_n$.

Thus, a **valuation** $V$ maps pairs $(g, D)$ to some real number denoted $V(g(X), D)$. For concreteness, a risk neutral Bayesian DM with model $(\pi, \Theta)$ has valuation $V(g(D)) = E_{\pi|D}(E_{F^{\theta^N}}(g(X)|\theta))$.

I now define what it means for a DM to have $(\pi, \Theta)$-conservative uncertainty valuation.

**Definition 1.** Fix a Bayesian model $(\pi, \Theta)$. A decisionmaker has $(\pi, \Theta)$-**conservative uncertainty evaluation** (CUE) if her preferences are represented by a function $V$ satisfying:

1. **(B1) $(\pi, \Theta)$-Conservatism:** $V(g(X), D) \leq E(g(X)|D) = E_{\pi|D}(E_{F^{\theta^N}}(g(X)|\theta))$, with equality attained if and only if $g$ is constant with respect to $D$.

2. **(B2) Consistency:** $V(g(X), D) \xrightarrow{P} E_F(g(X)) = \int g(x) dF(x)$ as $|D| \to \infty$.

---

4It is irrelevant that this data generation process is correctly specified: nature may simply select some unknown $F$. What is required is that in a game, all agents share this misspecified subjective view of data generation – that is, they have the same prior.

5As stated, this formulation accommodates only one source of uncertainty at each step. We will adopt the intuitive generalization, in the case of multiple independent sources, in later sections.
(B3) Continuity: \( g \mapsto V(g(X), D) \) is continuous.\(^6\)

(B4) Monotonicity: if \( g_1(x) \geq g_2(x) \) for all \( x \in X \) then \( V(g_1(X), D) \geq V(g_2(X), D) \).

Remark 1. A standard risk averse \((\pi, \Theta)-\)Bayesian is not \((\pi, \Theta)-\)CUE. Such a DM puts \( V(g(X), D) = u^{-1}\left( E_{\pi|D}(E_{F_{\theta}}(u(g(X)|\theta))) \right) \) and while this fails to violate B1, it violates B2: even in the limit where \( \pi \) is a singleton, Jensen’s inequality implies that \( u^{-1}(E(u(g(X)|\theta))) < E(g(X)|\theta) \) by strict concavity of \( u \), so the convergence in probability cannot hold.

Remark 2. As such, this measure is meant to capture a DM who has no aversion to a known risk, but does have such aversion to Knightian uncertainty (an unknown risk). Consistency means that as this uncertainty decays through data collection, the DM approaches the behavior of a risk neutral Bayesian DM. Conservatism enforces that this approach is always from “below.”

Remark 3. Many models of ambiguity aversion, interpreted correctly and specialized to the case of risk neutrality, feature decisionmakers with static \((\pi, \Theta)-\)CUE preferences:

1. A decisionmaker with smooth ambiguity aversion (Klibanoff et al., 2005), i.e. preferences represented by

\[
V(g(X), D) = \phi^{-1}\left( \int_{\Theta} \phi \left( \int_X g(x)dF^\theta(x) \right) d\pi(\theta|D) \right)
\]

for some strictly concave \( \phi \) is \((\pi, \Theta)-\)CUE.

2. A decisionmaker who maximizes the \( \alpha/2 \) quantile of the distribution of \( E(g(X)|\theta) \) under \( \pi|D \)\(^7\) or (to avoid this quantile possibly exceeding the mean and violating B1) takes the lower bound of some choice of \( 1-\alpha \) confidence set for \( E(g(X)|\theta) \) with \( E_{\pi|D}(E(g(X)|\theta)) \) strictly interior. The former is a generalization of the quantile maximization model (Rostek, 2010) to an Anscombe-Aumann setting.: Rostek considers preferences only on \( \Delta(X) \), whereas we are describing preferences on \( \Delta(X)^S \) satisfying risk neutrality with respect to the lottery and quantile maximization with respect to uncertainty.

3. A DM with maxmin expected utility (Gilboa and Schmeidler, 1989) under “full Bayesian updating” can be realized as \((\pi, \Theta)-\)CUE as well. For prior(s) \( \pi \in \Pi \) and likelihood

---

\(^6\)Equip the space of bounded functions \( g \) on \( X \) with the sup norm, regard it as a metric space with Chebyshev metric. Continuity is with respect to the metric topology.

\(^7\)Guaranteeing conservatism requires further assumption on \((\pi, \Theta, \{F_\theta\})\). We also modify (B1) to “if” instead of “if and only if.”
functions $f_\theta$ (see Epstein and Schneider (2003) for a refresher on notation), an MEU DM puts

$$V(g(X), D) = \min_{\pi \in \Pi} \int_{\Theta} \left( \int_X g(x) dF(x) \right) d\pi(\theta|D).$$

For any measure $\mu$ on the set $\Pi$ of possible priors, define $\pi^{Av}(\theta) = \int_{\Pi} \pi(\theta)d\mu$. Then DM is trivially $(\pi^{Av}, \Theta)$-CUE.

4. Hansen-Sargent constraint preferences (Hansen et al., 2006), where a worst case is taken over likelihood functions in a relative entropy ball around the “best-guess” likelihood (itself obtained via a form of Bayesian updating), is not $(\pi, \Theta)$-CUE for any $(\pi, \Theta)$ because the radius of the relative entropy ball stays fixed even as the best guess converges through learning: (B2) is violated. Related econometric methods fill that gap by picking $\varepsilon_n$ to define a likelihood ratio test varying with sample size and these approaches may in turn be $(\pi, \Theta)$-CUE.

2.1.2 Dynamic CUE

Let $t \in Z_{>0}$ index time and $D_t$ denote the data that DM has available at time $t$. For what follows, it is sufficient to define a decisionmaker’s cardinal evaluation of a finite sequence of real-valued functions of i.i.d. random variables

$$(g_t(X_t, D_t), g_{t+1}(X_{t+1}, D_{t+1}), g_{t+2}(X_{t+2}, D_{t+2}), \ldots, g_T(X_T, D_T)), \tag{1}$$

i.e. finite payoff streams received over time, given her data $D_t$. Denote this evaluation by $W_t$.

**Remark 4.** In full generality, defining a dynamic decision problem requires formalism for the actions $a_t$ that the decisionmaker can take in each period on which the functions $g_t$ and the data generation process may depend. Then, for example, $g_t$ is not only a function of $X_t$ and $D_t$ but also a function of the entire history of actions $\{a_s\}_{1 \leq s \leq t-1}$ and the action $a_t$ taken at time $t$. The action taken at time $t$ induces a continuation problem at $t + 1$ and beyond (depending on the horizon of the problem). An agent solves her finite horizon decision problem by maximizing, from the time $t$ (today’s) perspective, evaluation of the sequence of functions induced by all feasible sequences of actions she takes. My focus is narrower.

---

8Notice that we restrict to one new source of uncertainty per period and a very specific form for the data generated (and collected in datasets) over time.
Fix \( D_1 \neq \emptyset \) and let \( D_t = D_{t-1} \cup S_{t-1} \) for \( S_{t-1} \in \{ \emptyset, \{ X_{t-1} \} \} \)\(^9\) and \( t \geq 2 \). Notice that \( X_t \) is realized in period \( t \) so \( X_t \notin D_t \).

**Definition 2.** Say a decisionmaker has \((\pi, \Theta)\)-**dynamically conservative uncertainty evaluation** (DCUE) if her preference/evaluation over finite payoff streams admits a (recursive) representation given by

\[
W_T = V(g_T(X_T, D_T), D_T) \tag{2}
\]

and

\[
W_t = V(g_t(X_t, D_t) + \beta W_{t+1}, D_t), \tag{3}
\]

for \( 1 \leq t \leq T - 1 \), where \( V \) is a (static) evaluation satisfying (B1)-(B4) of Definition 1.

**Remark 5.** I will refer to the recursive representation in Equation (2) and Equation (3) jointly as (B5). To write (B5) in an alternative way, put \( W_T = g_T(X_T) \), define \( W_t = g_t(X_t) + \beta V(W_{t+1}, D_{t+1}) \) for \( 1 \leq t \leq T \). The decisionmaker’s objective is then to maximize \( W_1 = V(W_1, D_1) \).

**Remark 6.** Definition 2 specifies **recursive preferences**, which reduce evaluation and decision problems to a series of two period decision problems and deliberately imposes “dynamic consistency.” To do so requires defining a time aggregator and an certainty equivalent function in a recursive value equation: in this case, I require time aggregation to be additive with discounting and the certainty equivalent function to be the static conservative uncertainty valuation \( V \) developed in Definition 1.

### 2.1.3 Discussion of Recursivity

The evaluation of future payoffs (i.e. functions \( g_s(\cdot) \) for \( s \geq t + 1 \)) depends on information unknown (and uncertain) to DM at time \( t \). Even if DM can collapse the “continuation sequence” of payoffs into some certainty equivalent conditional on then-available data, from today’s perspective that certainty equivalent will still be a function of hitherto unknown data and thus possess uncertainty.

\(^9\)This formulation allows for the possibility that the DM does not learn anything. As a default, we will assume that DM always learns \( X_{t-1} \) going into period \( t \), prior to decisionmaking. This avoids dealing with multiple simultaneous sources of uncertainty that may arise (e.g. adverse selection).
Should this “knock-on” uncertainty be penalized in today’s valuation over and above the penalty it incurs in the future? This depends on the reason that uncertainty incurs a penalty – be it a fundamental aspect of decisionmaking or arising from some constraint. An insurer subject to period-by-period liquid cash constraints depending on his uncertainty about per-period portfolio payouts may not care to penalize uncertainty incurred in future periods over and above the expected penalty at that time: the penalty levied at time $t$ is only a function of the time $t$ uncertainty or cash flow, not future times’ or the stream’s. This may result in a (still dynamically consistent) evaluation like $W_1 = V(g_1(X_1), D_1) + \beta E(V(g_2(X_2), D_2)|D_1) + \ldots$ which does not have a recursive formulation as above, but in some ways is more orthodox.\textsuperscript{10} If this per-period constraint instead depends on equity – say in a capital adequacy ratio – then future uncertainty capitalized in the equity value should be penalized in this period, making a recursive penalization (as in our model) more appropriate.

Furthermore, the literature on dynamic decisionmaking has largely adopted recursive preferences in favor of dynamic consistency, unless its violations are explicitly the object of study. As such, dynamic generalizations of the decision-theoretic models in Section 2.1.1 are $(\pi, \Theta)$-dynamically conservative uncertainty evaluations. These include dynamic smooth ambiguity aversion (Klibanoff et al., 2009), (modified as earlier) dynamic quantile preferences (de Castro and Galvao, 2019, 2018), and recursive multiple priors models (Epstein and Schneider, 2007).

### 2.2 General Dynamic Monopolist Insurer Problem

This section shows that a risk-neutral monopolist insurer with conservative uncertainty evaluation may not be able to insure a risk averse (Anscombe-Aumann) expected utility buyer: this means that uncertainty can cause a prospective insurance market to fail in the presence of arbitrarily small per-period fixed cost, even when the dynamic incentive for exploratory underwriting (for data acquisition) is taken into account. Having established this, I show that sufficient exogenous data provision, through reducing the penalty associated with uncertainty, can bring these markets into permanent existence.

\textsuperscript{10}There’s a separate issue of why the constraint needs to – or not – penalize uncertainty only in given period profits (or cash), and to how or not this is tantamount to internalizing an uncertainty cost during the period (e.g. via Lagrange multiplier).
2.2.1 Game

Define primitives as in the previous section. Let $\mathcal{X} = [0, M]$ and regard $X \in \mathcal{X}$ as a “damage,” whereupon bearing the damage costs the agent. Before the game, Nature generates an infinite sequence of iid observations $Y_1, Y_2, \ldots, Y_s, X_1, X_2, \ldots$ from $F$: here $s$ denotes the cardinality of the “starter” dataset with which the consumers and the insurer are equipped. Index time by $t \in \{1, 2, \ldots, T\}$. There are two players, an insurer and a single buyer. In period $t$, the buyer faces a hitherto unknown damage $X_t$. Fix some nontrivial $(\pi, \Theta)$ Bayesian model for data generation. The insurer is $(\pi, \Theta)$-DCUE and the buyer is $(\pi, \Theta)$-Bayesian with mean-variance utility (and risk aversion $\gamma$).

At the beginning of each period, the insurer may offer a single contract and price $(B_t(x), p_t)$ to the buyer as a take-it-or-leave-it offer. Here $B_t(x)$ specifies the indemnity if loss $x$ occurs and satisfies

$$0 \leq B(x) \leq x \text{ for all } x \in \mathcal{X}, \quad 0 \leq B(y) - B(x) \leq y - x^{11} \text{ for } y \geq x$$

and $B$ bounded. Denote by $\mathcal{B}$ the set of functions satisfying the above conditions.

If underwriting occurs, the insurer incurs a writing cost $c$. When loss $X_t$ occurs, the insurer pays the customer the indemnity $B_t(X_t)$ per the contract. I assume no commitment or long-term contracts. There is common knowledge of $(\pi, \Theta)$ and preferences,\footnote{In particular, the buyer’s decision in each period is trivial and there is no ratchet effect.} and common information with $D_1 = \{Y_1, \ldots, Y_s\}$ and $D_{t+1} = D_t \cup \{X_t\}$ if and only if underwriting occurs in period $t$, else $D_{t+1} = D_t$. In particular, this means that the customer does not collect the data of her losses and update her beliefs if insurance is not underwritten, shutting down a prospective source of adverse selection.

I now set up optimizations explicitly. For convenience write $\rho = \pi|D$. Facing an unknown damage $X$, the uninsured buyer has expected utility

$$-E_{\rho}(E(X|\theta)) - \gamma \text{Var}(X) = -E_{\rho}(E(X|\theta)) - \gamma (E_{\rho}(\text{Var}(X|\theta)) + \text{Var}_{\rho}(E(X|\theta)))$$

Therefore the buyer does penalize uncertainty (the last term) and value the reduction of uncertainty that insurance performs: she collapses it into more risk. The buyer’s per-period value for insurance policy $B(X)$, say in period $t$, is thus

\footnote{This can be seen as a moral hazard condition, preventing the insured from willfully creating damage.}
\[ \Psi_{B_t} = E(B(X)|D_t) + \gamma (\text{Var}(X|D_t) - \text{Var}(X - B(X)|D_t)) \]

where expectations and variance are “total.”

Using the notation of Equation (3) and Equation (2), the insurer contemplates the sequence of contract functions \( \{B_t(\cdot,D_t)\}_{t=1}^{T} \) that maximizes \( W_1 \) and formulates a strategy involving offering a contract and price in each period. The insurer’s problem features trivial feedback: the decision to underwrite in period \( t \) only affects future periods through providing the insurer additional information: it does not directly constrain her action set in any way, and does not directly change what she can charge for a given contract. Coupled with the take-it-or-leave-it bargaining, this implies that the insurer extracts all surplus by pricing at \( \Psi_{B_t,t} \) when offering a contract \( B_t \) in period \( t \).

\( W_1 \) is thus given recursively by

\[
W_T = V((-c + \Psi_{B_T,T} - B_T(X_T))U_T, D_T) \\
W_t = V((-c + \Psi_{B_t,t} - B_t(X_t))U_t + \beta W_{t+1}, D_t)
\]

for \( 1 \leq t \leq T - 1 \) and \( U_t \) an indicator for underwriting at time \( t \), and the insurer’s problem, given \( D_1 \), is

\[
\max_{\{B_t(\cdot,D_t)\}_{t=1}^{T} \in \mathcal{B}} W_1.
\]

### 2.2.2 Discussion of Assumptions

My key assumption is to model insurers as having conservative uncertainty evaluation. Unlike consumers who deal with only one claim, insurers deal with several. Their modeling error or uncertainty – which is large in new or nonstationary markets with data paucity – is non-diversifiable across claims.\(^\text{13}\) As such, while insurers possess a diversification technology – modelled in reduced form through their risk neutrality – for known risks, they are still averse to uncertainty. Insurers’ penalization of this uncertainty may stem from risk aversion, ambiguity attitudes (Kunreuther et al., 1993), or financial constraints (e.g. a conditional tail expectation/value-at-risk or capital adequacy ratio constraint). If insurers are sufficiently sensitive to uncertainty, they may restrict or eliminate insurance coverage.

\(^\text{13}\)In other words, it behaves like the “aggregation risk” of earthquakes or cloud provider outages that blunts diversifiability across claims and challenges insurability.
While my model is written with respect to distributional uncertainty about one risk, arising from data paucity and mitigated by learning, it can also be interpreted as correlational uncertainty about a portfolio of identically distributed claims – i.e. uncertainty about the efficacy of diversification technology, which translates into higher expected costs of insurance.

The CUE assumption has other content. First, it means that insurers have an incentive to perform exploratory underwriting: loss-making underwriting in the present allows data collection that reduces uncertainty and increases future profits. This is consistent with insurers’ stated rationale for underwriting in certain new markets. Second, CUE can produce both intensive and extensive margin reductions of insurance coverage, which means that it has testable implications vis-a-vis market takeup rates and coverage limits.

Other assumptions are less central. Mean-variance utility is assumed to simplify this section: similar results apply with concave utility, control over the maximum index of Arrow-Pratt risk aversion, and the observation that convergence in probability implies convergence of the characteristic function. I abstract from moral hazard and adverse selection to show the isolated impact of CUE – although, for example, it is implausible that customers have detailed distributional data that they use to compute their valuations. I assume that data is only recorded and known when underwriting occurs: this assumption avoids endogenous creation of adverse selection, which only exacerbates the market failure problem.\textsuperscript{14} It also helps motivate introducing a fixed cost to underwriting in each period, which one may understand as a cost of verification; this avoids theoretical issues associated with censored disclosure and incentive compatibility,\textsuperscript{15} and more importantly, allows for a dynamic data trap. Otherwise, costless underwriting of trivial contracts (i.e. pure data collection) leads to sufficient data accumulation and forces a market into existence. Finally, the moral hazard assumption guaranteeing $B'(x) \leq 1$ seems innocuous but makes the space of contracts Lipschitz and facilitates the repeated application of Berge’s theorem.

\subsection*{2.2.3 Results}

The first result formalizes the notion of a \textbf{data trap}. Specifically, with an arbitrarily small fixed cost of underwriting and some bounded-below level of risk aversion, the insurer may optimally choose never to underwrite. This situation exists in spite of the exploratory incentives for underwriting vis a vis diminution of future uncertainty.

\textsuperscript{14}It also agrees with the observation that many firms are unaware of their own cyber-compromise and have no systematic procedures for data collection until they are insured.

\textsuperscript{15}While theoretically rich, a disclosure game with meaningful strategies seems entirely hypothetical.
**Theorem 2.1.** Let $\sup_{\theta \in \Theta} \text{Var}(F^\theta) = \overline{S}$. For any fixed cost $c > 0$ of underwriting, for any $\beta > 0$, define $\gamma(c)$ such that $\gamma(c) \overline{S} = c$. There exists some $\gamma > \gamma(c)$ such that the insurer with conservative uncertainty evaluation optimally never underwrites insurance.

*Proof.* See Appendix A. \hfill \Box

**Remark 7.** Altering the informational assumption – whereby the customer learns *whether or not* insurance is underwritten – does not change the result. It simply introduces adverse selection that leaves the insurer worse off. That a prospective customer cannot collect loss-relevant data if she does not purchase insurance is a strong assumption. A rationalization is to regard the customers as short-lived agents who are homogeneous vis-a-vis their preferences and whose only source of information is their (unmodeled) brokers, who update information only if contracts get underwritten.

**Remark 8.** Stronger statements hold for specific choices of DCUE preferences. For example, fixing $\gamma > 0$, it may always be possible to find specifications of $\{F^\theta(x)\}, (\pi, \Theta), \text{data, and potentially other preference parameters such that no contracting occurs in the static setting of Lemma A.1, even without a fixed cost:}$ the continuity argument allows for general functional forms but necessitates a cost.

Given the notion of persistent insurance market “failure” caused by CUE, can enough data facilitate exit of such a data trap? That is, can the government *jumpstart* a market? The second result states that a large enough dataset can be picked to jumpstart such an insurance market with arbitrarily high probability.

**Theorem 2.2.** Fix any $\varepsilon > 0$ and fixed cost $c > 0$. Let

$$\inf_{\theta \in \Theta} \text{Var}(F^\theta) > 0.$$  

Then there exists $\gamma > 0$ and starter data size $s = |D_1|$ sufficiently large such that the probability that the insurer optimally underwrites in every subsequent period exceeds $1 - \varepsilon$.

*Proof.* See Appendix A.2. \hfill \Box

**Remark 9.** Whether the $\gamma$-regions described in Theorem 2.2 and Theorem 2.1 overlap depends on the size of fixed costs and the minimum variance that $F$ may have, as codetermined by $(\pi, \Theta)$. Trivially, so long as a market would exist with a (arbitrarily close to) risk-neutral insurer, jumpstarting can be seen as a remedy to the failed market – albeit with a data requirement.
There thus exists some large enough starter dataset such that if the monopolist (and customer\textsuperscript{16}) were endowed with it at time zero, she would reliably continue underwriting in perpetuity. It is not, however, the case that getting the insurer to underwrite just once solves the nonexistence of the market in perpetuity: this requires stronger properties on valuation $V$ and implies a notion of “getting over a data trap.”

3 Algorithmic CUE and Applications

In this section, I (1) define an algorithmic or model-independent notion of conservative uncertainty evaluation, (2) introduce the “conservative confidence bound” (CCB) model as an example of such preferences, (3) solve for the optimal static contract in this model, and (4) offer comparative statics, including on the value of information and the probability of failure.

3.1 Algorithmic Conservative Uncertainty Evaluation

3.1.1 Definition

In the setting of Section 2, suppose that $\Theta$ indexes the set of all CDFs $\mathcal{F}(\mathcal{X})$ on $\mathcal{X}$ (which, for what follows, can be $\mathbb{R}$). Let $\pi_s = \text{DP}(s, H_0)$ be a Dirichlet Process prior (Ferguson, 1973) for some $H_0$ with full support and $s > 0$: then a draw $F^\theta \sim \pi_s$ defines a probability measure (distribution function) on $\mathcal{X}$ such that its evaluation at any partition of $\mathcal{X}$ is distributed according to a Dirichlet distribution. For $X_1, \ldots, X_n | F^\theta \overset{\text{iid}}{\sim} F^\theta$, the conjugate property of Bayesian updating with Dirichlet priors gives

$$\pi_s | X_1, \ldots, X_n \sim \text{DP} \left( n + s, \frac{s}{s+n} H_0 + \frac{n}{s+n} \sum_{i=1}^n \delta_{X_i} \right).$$

A decisionmaker who is $(\pi_s, \Theta)$-CUE thus satisfies

$$V(g(X), D) \leq \frac{s}{s+n} \int_{\mathcal{X}} g(x) dH_0(x) + \frac{n}{s+n} \left( \frac{1}{n} \sum_{i=1}^n g(X_i) \right).$$

Taking $s \to 0$, as in the Bayesian bootstrap (Rubin, 1981) and requiring $|D| > 0$, observe that a decisionmaker who is $(\pi_0, \Theta)$-CUE satisfies a special version of B1.

\textsuperscript{16}Relaxing customer rationality or concern with adverse selection would remove the need for this assumption, which may be realistic in a world with competition.
Notice there is no explicit prior-dependence to the RHS, so it serves as a nonparametric notion of conservative uncertainty evaluation to take to the study of algorithmic decision-making.

**Definition 3.** A decisionmaker has **algorithmic CUE** if her preferences are represented by a valuation \( V \) satisfying (B1\( ^* \)), (B2)-(B4). A decisionmaker has **dynamically algorithmic CUE** if her preferences are represented by a valuation \( V \) satisfying (B1\( ^* \)), (B2)-(B5).

Motivated by our usage of a mean-variance buyer in Section 2.2, I can compute the benchmark Bayesian posterior moments

\[
E(g(X)|D) = E(E(g(X)|D, F^0)|D) = \frac{1}{n} \sum_{i=1}^{n} g(X_i) = \hat{E}(g(X))
\]

and

\[
\text{Var}(g(X)|D) = E(g(X)^2|D) - E(g(X)|D)^2 = \frac{1}{n} \sum_{i=1}^{n} g(X_i)^2 - \left( \frac{1}{n} \sum_{i=1}^{n} g(X_i) \right)^2 = \hat{\text{Var}}(g(X))
\]

where I have implicitly switched expectation with integration and used properties of the Dirichlet process, so a \((\pi_0, \Theta)\)-CUE mean-variance decisionmaker simply has “empirical mean-variance” preferences. Such a decisionmaker is also prior-free.

### 3.1.2 CCB Valuation

I now define and introduce CCB valuation as an example of a algorithmic CUE preference. Fix \( \mathcal{X} = \mathbb{R} \). Given a dataset \( D \), define the empirical CDF \( F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x) \). The DKW inequality implies that for \( \alpha \in (0, 1) \),

\[
\varepsilon_n^2 = \frac{\log(\frac{2}{\alpha})}{2n},
\]

\[
L_n(x) = \max\{F_n(x) - \varepsilon_n, 0\}, \quad U_n(x) = \min\{F_n(x) + \varepsilon_n, 1\}
\]

form a uniform \((1 - \alpha)\) confidence interval for some true data generating \( F_c \). That is, for any CDF \( F_c \) and all \( n \),
Figure 1: Left: Plot of the nonparametric DKW \((1 - \alpha)\)-envelope \(\mathcal{C}_\alpha(D)\) for iid samples from an \(\text{Exp}(1/146)\) distribution. Right: Integration against the red CDF (if the data are presumed bounded) produces the Anderson upper bound for \(E(X)\) from Remark 10.

\[
P(\text{for all } x \in \mathbb{R}, L_n(x) \leq F_n(x) \leq U_n(x)) \geq 1 - \alpha.
\]

**Remark 10.** Per the observation of Anderson (1969), this construction allows the construction of conservative confidence intervals for the expectation of a bounded function of a random variable \(X \sim F\) given a dataset of observations. For certain functions, these expressions can be analytically computed (Romano and Wolf, 2002).

Set

\[
\mathcal{C}_\alpha(D) = \{ G : \mathbb{R} \rightarrow [0, 1] \text{ increasing}; L_n(x) \leq G(x) \leq U_n(x) \}.
\]

**Definition 4.** For a bounded function \(M\), independent draw \(X \sim F\), and dataset \(D\), say a decisionmaker has **conservative confidence bound (CCB) valuation** if her preferences are represented by

\[
V(M(X), D) = \min_{G \in \mathcal{C}_\alpha(D)} \left( \int M(x) dG(x) \right).
\]

A CCB decisionmaker is also algorithmic CUE. First, \(V(M(X), D) \leq \int M(x) dF_n(x)\), implying conservatism. Consistency is implied by the shrinking DKW envelope width \(\varepsilon_n\) along with the strong convergence of the empirical mean. Monotonicity and continuity are inherited. We may define a dynamic CCB decisionmaker in the obvious way, extending the above definition recursively. In moving from a static to dynamic problem, the choice of worst-case
measure may complexify substantially.

**Remark 11.** Relative to a decisionmaker who computes the bottom endpoint of a \((1 - \alpha)\)-posterior credible interval for \(E(M(X)|D)\) using a Bayesian bootstrap procedure (Rubin, 1981) – as a nonparametric twist on a previously discussed augmentation of the quantile maximization model\(^17\) – a CCB decisionmaker is typically more conservative. Not all CDFs inside the DKW envelope are equally likely given the data, and a Bayesian method may leverage this fact whereas CCB evaluation will not. Nonetheless, there is a sense in which the scale of CCB estimate's conservatism – given that it is based on a nonparametric confidence interval – is optimal in \(n\) (see Romano and Wolf (2000)), justifying the approach.

### 3.1.3 IDP Valuation

An even more tractable algorithmic CUE preference can be generated – in strong analogy to MEU with full Bayesian updating – using the imprecise Dirichlet prior formulation of Benavoli et al. (2015). This serves as an alternative to the \(s \to 0\) assumption of the Bayesian bootstrap. Specifically, let \(\mathcal{F}\) be the set of all probability measures on \(\mathcal{X}\) and let, for fixed \(s > 0\),

\[
\mathcal{H} = \{\text{DP}(s, H_0) : H_0 \in \mathcal{F}\}
\]

be the set of all Dirichlet processes, letting \(H_0\) span \(\mathcal{F}\). To perform Bayesian updating of this set of priors given observations \(D = \{X_1, \ldots, X_n\}\), we update prior by prior, exploiting the conjugacy of the Dirichlet process. Then, for any bounded function (else we get value \(\pm \infty\)) we may define upper and lower expectations over the set of posteriors. Use the lower expectation to define a valuation.

**Definition 5.** For a bounded function \(M\), independent draw \(X \sim F\), and dataset \(F\), say a decisionmaker has **IDP valuation** if her preferences are represented by

\[
V(M(X), D) = \frac{s}{s+n} \inf_{x \in \mathcal{X}} M(x) + \frac{n}{s+n} \left( \frac{1}{n} \sum_{i=1}^{n} M(X_i) \right).
\]

Observe that an IDP decisionmaker is also algorithmic CUE.\(^18\)

\(^{17}\)See also Tindemans and Strbac (2017) for usage of this method in an unrelated setting.

\(^{18}\)Unfortunately, \(s\) no longer has the advantage of corresponding to a confidence level as \(\alpha\) does in CCB preferences.
3.2 Static CCB Monopolist Insurer Problem

In this section I characterize the optimal single-period contract that a CCB monopolist offers. I offer comparative statics and discuss the value of government intervention.

3.2.1 Insurer’s Problem

Consider the setting of Section 2.2.1 with \( X = \mathbb{R}^+ \) and \( T = 1 \), i.e. a static contracting problem. Let the insurer have CCB preferences for some \( \alpha \) small (corresponding to the significance level for a statistical test), and let the buyer have empirical mean variance preferences; thus the insurer is algorithmic CUE and the buyer is \((\pi_0, \Theta)\)-CUE mean variance as in the Section 3.1.1. Set underwriting cost \( c = 0 \).

With notation as in Section 2.2.1, the insurer offers a contract \( B \) she can price at

\[
\psi_B = \hat{E}(B(X)) + \gamma(\text{Var}(X) - \text{Var}(X - B(X)))
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} B(X_i) + \gamma \left( \text{Var}(X) - \frac{1}{n} \sum_{i=1}^{n} (X_i - B(X_i) - (X - \bar{X}))^2 \right)
\]

where \( \bar{A} = \frac{1}{n} \sum_{i=1}^{n} A_i \) as usual. The cost to a FAA insurer of underwriting a (bounded) policy \( B \) is the worst integral against the \((1 - \alpha)\) DKW envelope, obtained by integrating along the bottom. For \( \varepsilon_n^2 = \log(2/\alpha)/(2n) \) and \( k = [n\varepsilon_n] \) this is

\[
C_B = 0 + B(X_{(k+1)}) \left( \frac{k + 1}{n} - \varepsilon_n \right) + \frac{1}{n} \sum_{j=k+2}^{n} B(X_{(j)}) + \varepsilon_n B(X_{(n)})
\]

which reassigns all mass from the first \( k \) points and some from the \( k + 1 \)-st to the maximum observation (see Figure 1). That is, the first \( \varepsilon_n \) worth of mass in the distribution gets moved from the lowest values to the highest value.

The insurer seeks to maximize, in \( B, \psi_B - C_B \), provided it exceeds zero \( (c, \text{in a model with fixed costs}) \).

**Remark 12.** I note some simplifications to the space of possible contracts:

1. Because the buyer has no value for coverage offered for losses above \( X_{(n)} \) it is immediate that we may set \( B(t) = B(X_{(n)}) \) for all \( t > X_{(n)} \). Even if this weren’t the case and the buyer had strictly increasing value for coverage, \( B \) would have to be bounded, else \( C_B \) would be infinite.
2. Recall the requirements that (i) $B$ is weakly increasing, (ii) $0 \leq B(x) \leq x$, and (iii) $B(y) - B(x) \leq y - x$ for all $y \geq x$.

3. As stated, we need only define $B$ on the values $\{X_i\}$: neither does the insured derive value from the intermediate segments, nor do these enter into $C_B$. This has two consequences:

(a) Optimal contract(s) are generally not unique.

(b) In extending $B$ to intermediate values (sometimes constraints, if the slope is 0 or 1, will imply only one extension), it is without loss of generality that for some optimal $B$, the slope, where defined, is always either 0 or 1.

Of course, these intermediate values have legal relevance vis-a-vis the interpretation of the contract; otherwise, there is the question of interpolation method.

4. The space of contracts $B$ is thus compact so (at least) the results of Section 2.2.3 hold.

Reindex $\{X_i\}$ so that $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$. Then the insurer’s problem may be restated as

$$
\max_{\{b_1, \ldots, b_n\}} \left( -\varepsilon_n b_n + \frac{1}{n} \sum_{j=1}^{n} b_j + \left( \varepsilon_n - \frac{k}{n} \right) b_{k+1} + \frac{\gamma}{n} \left( n \widehat{\text{Var}}(X) - \sum_{i=1}^{n} (x_i - b_i - (\bar{x} - \bar{b}))^2 \right) \right)
$$

subject to

$$
(\lambda) : \quad 0 \leq b_1 \leq b_2 \leq \ldots \leq b_n \\
(\mu) : \quad b_i \leq x_i \\
(\nu) : \quad (b_{i+1} - b_i)/(x_{i+1} - x_i) \leq 1
$$

3.2.2 Optimal Contract

In this subsection, I characterize an optimal static contract the CCB insurer offers. Given any permissible contract $B : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$, call an interval $[a, b]$ a flat region (F) for $B$ if $B'(x) = 0$ for all $x \in (a, b)$. Similarly, call an interval a slant region (S) for $B$ if $B'(x) = 1$ for all $x \in (a, b)$. By Remark 12, some contract $B$ partitioning $\mathbb{R}^+$ into flat and slant regions solves the CCB insurer’s problem. I refer to a finite sequence of alternating flat and slant regions induced by $B$ as a partition of $\mathbb{R}_{\geq 0}$ as a configuration. With this language:
Theorem 3.1. Any solution to the CCB insurer’s static contracting problem features the same loss limit less than $\max_i x_i$. There exists a solution $B$ partitioning $\mathbb{R}_{\geq 0}$ into one of the following configurations: $F, SF, FSF$, or $SFSF$.

Proof. See Appendix C.1.

Remark 13. Without loss of generality, this means (i) the optimal contract always features a loss limit, (ii) the optimal contract changes slope at most thrice and does not feature coinsurance, (iii) the optimal contract may feature a deductible, and (iii) the market may fail (the optimal contract is 0). See Figure 2 for the various configurations that arise.

The tradeoff between the risk premium the buyer is willing to pay and the ambiguity cost the insurer incurs governs the optimal contract. Because there are no frictions in this model, the contract is best understood in terms of its deviation from full insurance. In a problem without the monotonicity, slope, and indemnity constraints, this happens precisely where the Anderson $(1-\alpha)$-worst case measure deviates from the empirical measure.

Limits arise because for the insurer to control her expected loss at the $(1-\alpha)$ level, she must bound the amount she pays out (thus defining the worst-case measure). The size of this limit, at least in an unconstrained contract, equates the marginal risk premium with the (constant) marginal ambiguity cost.

Analogously, the insurer has strong incentives to offer coverage on $[x_1, x_k]$, because she perceives no cost. These drastically lessen at $x_{k+1}$ and $x_{k+2}$. Because the contract must be monotone, the insurer needs to flatten – perhaps completely – the contract in this region. Depending on the decrease in the incentives and the spacings between data values, this may manifest either in an intermediate flat region (i.e. the first F in a SFSF configuration) or a deductible. Both in reality (if ever seen) and in simulations, SFSF contracts are far rarer than FSF contracts.

3.2.3 Comparative Statics

In this section, I discuss the dependency of the optimal contract and market outcomes upon parameters $(\gamma, \alpha)$ and future data realizations. First, I define four relevant market outcomes. These are:

1. Market existence, an indicator function denoted by $I(D, \alpha, \gamma)$.

2. The effective policy limit, $L(D, \alpha, \gamma) = b_n(D)$. 

20
Figure 2: Pictured are three simulated optimal contracts. The simulations use $n = 1000$ samples from a $\text{Exp}(1/146)$ distribution, $n = 1000$ samples from a lognormal distribution with $\mu = 2.2$ and $\sigma = 2.4$ and $n = 300$ samples from a $U[0,8000]$ distribution. In all simulations, $\alpha = 0.01$, and $\gamma = 0.7 \times 10^{-3}$. Note that in the second figure, the segment $x_{n-2} \rightarrow x_{n-1}$ with slope less than one can be augmented to have slope 1 and then slope 0 as per Remark 12. As such, these examples respectively have configurations FSF, SF, and SFSF cited in Theorem 3.1.
3. Risk transfer

\[ \mathcal{R}(D, \alpha, \gamma) = \gamma \left( \widehat{\text{Var}}(X|D) - \widehat{\text{Var}}(X - B^*(X, D)|D) \right) \] :

absent fixed costs, this measures the insurer’s expected dollarized profits.

4. Insurer “perceived profit” (net of ambiguity cost): the value of her optimized objective function \( \Pi(D, \alpha, \gamma) \).

With respect to risk-aversion and the confidence threshold, the following result is immediate:

**Corollary 3.1.** The functions \( I, L, \mathcal{R}, \Pi \) all increase in both \( \alpha \) and \( \gamma \).

The intuition is clear: increasing \( \gamma \) increases risk aversion and the surplus a market could create (if extant), holding all other things fixed. Increasing \( \alpha \) reduces the insurer’s uncertainty aversion: in the formulas this shows up through \( \varepsilon_n \), making it smaller. This reduces her distortion vis-a-vis a “empirical” risk neutral insurer.

What happens to the optimal contract when a single datapoint is added to the dataset \( D \)? Adding a large datapoint above the abscissa of the limit (if the market exists) increases the limit either by increasing \( s \) or increasing the slope of the segment connecting \( b_s \) to \( b_{s+1} \). If a market does not exist, it helps the market satisfy the condition for existence. Adding a smaller datapoint, if the market exists, results in the limit decreasing (because the \( n - s \) points from the top remain unchanged and \( n \) has incremented to \( n + 1 \)) slightly though it may decrease the size of the intermediate flat region if one exists. If the market does not exist, it only helps the market exist if it is sufficiently large – which is determined by the other data. Figure 3 provides an illustration of how the optimal contract changes as data are successively appended to it.

How about the expected impacts on \( I, L, \mathcal{R}, \Pi \) of dataset expansions? Evaluating such an expectation would tell a distributionally informed government the private and social returns to government data collection or compilation. If the government is not so informed, it may need to construct an estimator (potentially with resampled observations from \( D \)) or bounds (e.g. using the DKW envelope) for this quantity. Even with knowledge of \( F \), a calculation requires quantitative control over the contract changes discussed above and is much stronger than consistency-based claims (e.g. \( I_k \xrightarrow{P} 1 \)). I make the following conjecture:

**Conjecture 3.1.** Fix \( D \). Define

\[ Q^G_k(D) = E_{(Z_1, \ldots, Z_k) \sim G}(Q(D \cup \{Z_1, \ldots, Z_k\}, \alpha, \gamma)) \] for \( Q \in \{ I, L, \mathcal{R}, \Pi \} \), i.e. the expected market quantity \( Q \) when \( k \) points are sampled iid from
Figure 3: Evolution of optimal contract given a dataset \( D \) with \( n = 4, 5, 6, 16, 17, 18 \) data points appended (row-by-row). The last appended point is depicted in red, old values of the optimal contract are greyed out, and the present optimal contract is shown in blue. Recall that segments with slope \(< 1\) can be replaced, without economic change, by segments with slope 1 and 0. In the language of Theorem 3.1, the first contract has configuration \( F \) and the rest are \( FSF \).
some distribution $G$ and appended to $D$. If $G$ first order stochastically dominates $F_n$, all of $\Pi^G_k(D)$, $I^G_k(D)$, $R^G_k(D)$ and $L^G_k(D)$ are monotone increasing in $k$.

Finally, regarding the impact of the amount of data upon market existence, I make the following conjecture:

**Conjecture 3.2.** Fix $F$. The unconditional probability of market existence in a market with $n$ datapoints $E_{D=(X_1,\ldots,X_n) \sim F} \cap (I_n) = \mathcal{I}_n(F)$ monotonically increases in $n$.

Figure 4 offers some limited evidence from simulation results. ?? 3.2 is supported by the general intuition of Theorem 2.2.

![Figure 4: Plot of $\mathcal{I}_n(F)$ with $m = 400$ simulations for each datapoint. Here, $F = \text{Expo}(1/146)$, $\gamma = 0.7 \times 10^{-3}$, and $\alpha = 0.01$.](image)

**3.2.4 Asymptotic Coverage**

While the effective policy limit $L^* = b_n$ and risk transfer can be computed for a fixed dataset $D$, the algorithm relies on condition-checking. Thus it is not easy to compute its expectation
over all datasets of a particular size, given a fixed distribution $F$. Nonetheless, we can provide asymptotic results that capture the qualitative impact of CUE. Evidently $F(L^*) \to 1$ as $n \to \infty$, but another notion of the amount of market coverage would be to consider $L^*/X_{(n)}$—i.e. compare how the current policy limit compares to the largest hitherto experienced loss.

Because the index

$$s = \max \left\{ t : \sum_{j=t+1}^{n} (x_j - x_t) \geq \frac{n\varepsilon_n}{2\gamma} \right\}$$

helps define an upper bound for the abscissa at which the optimal policy limit is obtained (if one is), it follows that (up to a constant depending on $\gamma$), the ratios $x_s/x_n$ and $x_{s+1}/x_n$ will determine the asymptotic behavior of $L^*/x_n$. I make the following loose conjecture:

**Conjecture 3.3.** $L^*/X_{(n)} \overset{P}{\to} 1$ only if the tail of $F$ is sufficiently fat. If the tail is sufficiently thin, $L^*/X_{(n)} \overset{P}{\to} 0$.

Appendix C.2 contains formal results partially characterizing $L^*/X_{(n)}$ as above. My strategy is to “rule out” possible locations of $L^*$ and then leverage results on the ratios and spacings of order statistics. To define and control tail behavior, I use the parameterized maximum domains of attraction in the Fisher-Tippett-Gnedenko theorem (see, e.g. Embrechts et al. (2013)).

Economically, for fat-tailed enough risks, data collection (if embarked upon at all) leads to asymptotically full insurance markets: demand from risk aversion wins out over insurers’ uncertainty. For thinner-tailed risks, this may not be the case: even in the limit of infinite data, loss limits may be much smaller than historically experienced losses. In this case, insurer uncertainty has a real bite relative to risk aversion. Intuitively, customers’ mean-variance preferences are very important for this result: their demand for insurance explodes when the risk distribution is sufficiently fat-tailed.

### 3.3 Discussion: Contract Features and Empirical Predictions

The optimal CCB monopolist contract matches several features of actual cyberinsurance contracts, like loss limits, (occasional) deductibles, and an absence of coinsurance (see, e.g. the analyses of contracts in Woods et al. (2021) and Romanosky et al. (2019)).\(^\text{19}\) In this section, I describe how these features might obtain in alternative models (involving classical

\(^\text{19}\) Coinsurance and sublimits have begun to appear with respect to ransomware coverage (Jones, 2021), but neither is this the norm nor is this conventional moral hazard. See Balasubramanian (2021) for a discussion of the *external adversarial hazards* present in that case.
frictions and constraints) and how one might compare them to the CCB model. I also discuss why these other stories are anecdotally implausible, and predictions of the CCB model with respect to longitudinal contract and loss data.

First and foremost, in a baseline Bayesian model without frictions, full insurance obtains. Data paucity simply increases risk transfer in the market. If an insurer is constrained only to offer one price and faces adverse selection about the customer’s mean damage (“type”), less data—insofar as it manifests as a larger damage variance conditional on type—leads to more insurance market coverage (on the extrinsic margin).\(^{20}\) If the insurer can offer a menu of contracts to screen customers, the policy limits (intrinsic margin coverage) she offers increase with data paucity.\(^{21}\) In this sense, simple models predict that data paucity expands insurance markets rather than contracting them.

Ignoring this comparative static with respect to data paucity, we may still try to reconcile equilibrium menus under adverse selection with observed cyberinsurance contracts. Analysis of monopoly contracts (see Chade and Schlee (2012) and Stiglitz (1977)) deals only with binary loss distributions, which cannot distinguish between coinsurance (i.e. nontrivial region with \(0 < I'(X) < 1\)) and policy limits. Nonetheless, under monopoly, higher risk types get fuller (with highest types getting full) insurance, while elsewhere there are downward distortions. Some low types may not get insurance at all.\(^{22}\) Results for the separating equilibrium in the variable loss severity case (with perfect competition, see Fluet and Pannequin (1997) and Young and Browne (1997)) are similar: high types get full coverage, while low types receive contracts with limits, coinsurance, and (sometimes) deductibles. These models have at least two observational issues. First, no cyberinsurer offers anything resembling full insurance to any “type:” the largest policy limits tend to be some 0.2%-5% of annual revenue (see Marsh (2020) and Cowbell Cyber (2020)), which is small both compared to other types of insurance and maximum possible losses. Second, in models with variable loss severity, coinsurance generically (and always does under standard assumptions like monotone likelihood ratio) occurs\(^ {23}\); this fails to occur in real contracts.

\(^{20}\) Of course, if data paucity is modeled as simply increasing the extent of adverse selection, then it would shrink the market.

\(^{21}\) This follows from augmenting the results of Young and Browne (1997).

\(^{22}\) Chade and Schlee (2012) suggest that risk premium (i.e. price minus actuarial cost) follows an inverse S pattern with respect to damage type (i.e. limit selection), which is a testable prediction with the right data.

\(^{23}\) It takes special cases or binding constraints for contracts not to have coinsurance. For example, if low and high types have identical severity distributions and merely have different probabilities of damage, then the optimal contract features a deductible (but no limit).
Models with moral hazard suffer similar issues in terms of observables. It is well-known (Winter, 1992) that models where costly customer effort affects only the probability of loss (resp. only the expected loss conditional on a loss happening) lead to deductibles (resp. full insurance until a point, at which coinsurance occurs); models featuring both will generally have strict coinsurance under MLRP and the usual assumptions for a first order approach. The absence of loss limits and the pervasive presence of coinsurance are hard to reconcile with observed market contracts, whose salient feature is low limits.

The argument that moral hazard and adverse selection are relatively unimportant in cyberinsurance has plenty of anecdotal support as well. As concerns adverse selection, few companies are even aware of their risk profile, and fewer yet have any quantitative conception of it (Marsh (2019) puts the number of firms with such a quantitative assessment at approximately 30%). Firms are generally unaware if they have been compromised and lean heavily on cyberinsurers for superior security knowledge. Symantec (2013) finds that 50% of small businesses have been compromised, while a survey of 1083 firms (Greig, 2019) finds significant underestimation of breach probability and severity. Regarding moral hazard, monitoring technology and security integration are one of the strong points of cyberinsurance; contractually mandated realtime threat detection along with software installations (e.g. tracking) and de-installations (e.g. of Windows Remote Desktop) are the norm in the industry. And the preponderance of attacks are not targeted retaliations; instead, they are diffusely “sprayed” through phishing, business email compromise, or software exploits.

The CCB model also delivers predictions for how contracts evolve in response to data arrival, reflecting the effects of learning. Loosely, as in Figure 3, small losses lead the loss limit to shrink slightly (by a diminishing amount as dataset size increases) while large-enough losses lead limits to jump. An adverse selection model with Bayesian learning would deliver a similar result if a large loss induces updating of variance, hence creating demand. This prediction, however, runs contrary to the data. Large losses typically lead insurers to reduce limits and raise prices, if not exit the market entirely, much more consistent with capital depletion (see Froot and O’Connell (2008) on catastrophe insurance). As such, bringing the CCB model to longitudinal contract and loss data would also require introducing capital market frictions – both slow-moving capital (Duffie, 2010) and insurer capital constraints. Capital constraints (e.g. capital adequacy ratios that stipulate a balance sheet must be

---

24 This has happened recently in the cyberinsurance market. See US Government Accountability Office (2021) for an overview.

25 In the 1960s, private US flood insurers, on account of large losses, exited the market en-masse. Today, some 95% of flood insurance policies are backed by the National Flood Insurance Program, a federal initiative.
able to absorb, with sufficient buffer, a $(1 - \alpha)$-quantile loss alone are insufficient to produce insurance limits without CUE; an insurer will simply underwrite more (independent) policies and diversify away from a binding constraint. In tandem with prudential margining and loss reserving, limited capital, non-pledgeability of premiums, and/or lender/reinsurer CUE, these constraints can produce limits and better match intertemporal patterns.

4 Data Policy

Earlier sections show how model uncertainty may impair insurance provision. The main determinant of this uncertainty is the (size of the) data sample $D$ that insurers possess. In this section, I discuss government investment in datasets and datasharing schemes as possible policy measures to foster insurance markets for new or nonstationary risks.

Why pursue these measures and how to do so? First, although we basically cannot reduce the correlatedness of (say) earthquake damages, we can invest in data through exploratory underwriting or auditing: data acquisition serves effectively as a diversification technology in this context. Second, as is oft-discussed, data is nonrivalrous and shareable across markets and firms. Of course, private incentives for data acquisition and sharing may differ from social incentives: appraisals of data value may differ, insurers may not capture all the gains from their data acquisition – whether it accrues to buyers or other insurers, and full sharing outcomes – even if efficient – may be hard\textsuperscript{26} to reach.

4.1 Kickstart Data Policy in CCB Monopolist Insurer Problem

The models above suggest (e.g. Theorem 2.2) that providing increased underwriting data might lead to increased risk transfer through the insurance market. In this subsection, through the CCB model, I contemplate the returns to government investment in such a dataset $J$ for insurers to use. Such data might, for example, be generated through detailed audits of prospective customers or merged from other disaggregated data the government has. To simplify analysis, I restrict to $T = 1$ and regard exploration as an insurer buying data at price $c$ prior to underwriting. Further, I assume that data is made available to buyers forming their expectations.\textsuperscript{27}

\textsuperscript{26}For example, neither government data collection nor mandated data sharing has taken root in cyber; voluntary sharing efforts (e.g. Information Sharing and Analysis Centers and contributory databases) have been extremely limited and localized thus far.

\textsuperscript{27}As noted earlier, this assumption is more innocuous in a model with perfect competition.
While the model of Section 3.2 has no inefficiency, there are at least three reasons a government data purchase differs from a firm’s. First, conservatism towards future data’s usefulness may differ: the government may calculate $Q^G_m(D)$ using some distribution $G$ different from (perhaps stochastically dominating) the insurer’s upper Anderson bound. Second, data collection costs may differ: I assume that government data collection is more expensive. Third, the insurer and the government may have different objective functions: while the insurer seeks to optimize her perceived profit (net of ambiguity cost), the government may care only to optimize risk transfer. For simplicity I shut down this third consideration and assume the government targets $\Pi^G_m(D)$. Accordingly, we have the following claim:

**Proposition 4.1.** Fix a CCB monopolist market $(D, \gamma, \alpha)$ and let $c = 0$. Then for some $m > 0$, there exist $(\eta_I, \eta_G)$ per-datapoint acquisition costs with $\eta_G > \eta_I$ such that the government – if evaluating $\Pi^F_m(D)$ according to the empirical measure $F_n$ – would buy a dataset of size $m$, and the monopolist CCB insurer would not.

**Proof.** See Appendix D.

**Remark 14.** This claim is most interesting when the market $(D, \gamma, \alpha, c = 0)$ is failed, which is possible in a CCB model. Even with higher-than-private data acquisition costs, differing uncertainty attitudes may mean the government finds it profitable to jumpstart a failed market.

The point is of course stronger if the insurer does not extract all surplus through pricing. With finer control over the marginal returns to data, we might have:

**Conjecture 4.1.** There exist failed markets $(D, \gamma, \alpha)$ and costs $\eta_G > \eta_I$ such that the insurer buys no dataset of any size whereas the government optimally would.

There are obviously caveats to practical interpretation of these results. The government should not buy data that the insurer would anyway: doing so wastes $m(\eta_I - \eta_G)$. A dynamic model might better formalize such crowding out. Furthermore, knowing or estimating the returns to data acquisition is challenging. The assumptions about government knowledge of $D$ and its uncertainty attitude may not fit various situations.

Nonetheless, in the context of failed insurance or lending markets, the government might consider acquiring, merging, or providing datasets to mitigate the substantial uncertainty the relevant insurer or lender faces. While this approach has been pursued with respect

---

28 In particular, the government does not know $F$, so the “omniscient” calculation of returns to data $Q^F_k(D)$ (conjectured monotone) is not feasible: $G$ must also be some estimator based on the data. Assuming that the government does not know $D$ – or only knows $|D|$ – complicates issues further.
to meteorological data and fire and flood maps (for property insurers’ use), it may also be appropriate for distributional information on individuals’ earnings by field of study, possibly by educational institution. Such information is not collected by most (typically not-for-profit) universities but would eliminate a considerable friction in jumpstarting the income-share agreement (ISA) market\(^ {29}\) and the expansion of thin file credit markets (Carroll and Rehmani, 2017) for recent graduates more generally.

### 4.2 Data Sharing

In this subsection, I suggest benefits to mandated data sharing between insurers. Simply put, when insurers in similar markets share (anonymized or distributionally binned) loss data, they may reduce their respective uncertainty and improve social surplus from expanded insurance provision. Requiring such sharing as a matter of policy may be necessary if private sharing does not reach the collusive optimum. Much as diversifiability of claims implies economies of scale in insurance, so does data collection through reducing model risk. This consideration may help inform conventional antitrust logic in new insurance markets.

While my main argument is colloquially understood, it requires CUE to work. As such, it does not feature in standard economic discussions of (typically non-anonymized) datasharing, which focus strongly on inferential aspects of adverse selection in such contexts as individual credit scoring by bureaus. Full analysis of datasharing is beyond the scope of this paper because it requires – whether within the CCB model or more generally – a formulation of competition; Appendix B lays out necessary ingredients. Instead, I use a reduced form, one-period Bertrand competition model with stylized assumptions (informed by earlier sections) to make my point.\(^ {30}\)

Consider similar but separate markets \(j = 1, 2\) and insurers \(i = 1, 2\), each associated primarily with one market (its namesake). Assume a single buyer exists in each market with known value \(v\). Denote by \(d_{i,j}\) the data that insurer \(i\) has about the risk in market \(j\). The data insurers possess determines their underwriting costs in each market; this cost \(c_{i,j} = c(d_{i,j}, d_{i,-j})\) decreases and is convex in each argument, i.e. \(\frac{\partial c_{i,j}}{\partial d_{i,k}} < 0\) and \(\frac{\partial^2 c_{i,j}}{\partial d_{i,k}^2} > 0\).

\(^{30}\)A dynamic model might explore how datasharing crowds out exploratory underwriting – a cost that does not appear in my model. Nonetheless, in a market with substantial adverse selection as well, the fact that insurers are only revealing anonymized data may allow them to protect their rents and for data sharing to still have a positive effect.

---

\(^{29}\)See Ritter and Webber (2019), who claim “ultimately, the future success of the ISA depends heavily on the quality and length of the institutional and supplemental regional and national data on earnings patterns” and suggest that a (hitherto unpursued) merging of data already possessed by the US Department of Education and the IRS could help accomplish precisely these goals.

\(^{30}\)A dynamic model might explore how datasharing crowds out exploratory underwriting – a cost that does not appear in my model. Nonetheless, in a market with substantial adverse selection as well, the fact that insurers are only revealing anonymized data may allow them to protect their rents and for data sharing to still have a positive effect.
for \( k \in \{ j, -j \} \). Assume, for simplicity, that \( c : \mathbb{R}^2 \rightarrow \mathbb{R} \) is differentiable everywhere. Data about market \( i \) is more useful in market \( i \) than data about market \( j \), i.e.

\[
\frac{\partial c_{i,j}}{\partial d_{i,j}} < \frac{\partial c_{i,j}}{\partial d_{i,-j}} < 0.
\]

Each insurer is endowed with a dataset of size \( n \) about her “associated market” and none about the other market:

\[
d_{1,1} = n, \quad d_{1,2} = 0 \quad \text{and} \quad d_{2,1} = 0, \quad d_{2,2} = n.
\]

These assumptions describe situations where the same type of risk is being written in geographically, demographically, or technologically (e.g. in terms of software providers) disparate markets where insurers have specialization on the basis of past experience, data collection, or chance. Inspired by CUE, anonymized or (solely) distributionally relevant data reduces underwriting costs. And while data from other markets is informative, it is trickier and noisier to translate on a per unit basis.

At the beginning of the period, each insurer elects to share fraction \( \alpha_i \in [0, 1] \) of its dataset (say rows selected uniformly at random) and fraction \( \gamma_i \in [0, 1] \) of its primary market profits with its competitor. Subsequent competition is Bertrand: after datasharing, each insurer \( i \) simultaneously submits bids \( \{p_{i,j} : j \in \{1, 2\}\} \) to both markets and the winner in each market gets \( (p_{i,j} - c_{i,j})I(v > p_{i,j}) \). Primary market profits are then shared per each party’s election. Finally, I assume

\[
c(n, n) < v < c(n, 0) < c(0, n).
\]

This means that insurers would not underwrite without any datasharing (e.g. their uncertainty cost is too high) but that full datasharing would imply the existence of insurance market surplus.

Proposition 4.2. In all symmetric subgame perfect equilibria of this game:

1. Each insurer shares some fraction \( \alpha \in [0, \bar{\alpha}] \) of its data with the other, for

\[
\bar{\alpha} = \sup\{\rho : c(\rho n, n) \leq v\}.
\]

2. Only insurer \( i \) (may) “enter” market \( 31 \). This occurs when \( \alpha \geq \alpha^* = \sup\{\rho : c(n, \rho n) \leq v\} \). Thus, when \( \alpha \geq \alpha^* \) insurer \( i \) charges \( p_{i,i} = v \) in its associated mar-

\[31\] That is, \( p_{-i,i} \geq c_{-i,i} \geq v \) with equality if and only if \( \alpha = \bar{\alpha} \)
Figure 5: Cost functions \( c(n, \alpha n) \) and \( c(\alpha n, n) \) of insurer \( i \) and insurer \( -i \) in market \( i \). It is not profitable for (even) insurer \( i \) to underwrite when \( \alpha_{-i} < \alpha^* \). Sharing any fraction \( \alpha_i < \tilde{\alpha} \) means insurer \( -i \) does not enter, allowing insurer \( i \) to price at \( v \). The red line depicts insurer \( i \)'s equilibrium pricing decision (when relevant) in the subgame where \( \alpha = \alpha_i = \alpha_{-i} \) is symmetrically chosen.

ket.

3. Neither firm shares primary market profits if these are nonzero: \( \gamma = 0 \) unless \( \alpha < \alpha^* \) in which case any \( \gamma \) can obtain in equilibrium.

Proof. See Appendix D.

I restrict my focus to symmetric equilibria, where even allowing profit-sharing, the amount of data insurers will possibly share with each other is bounded. This is not true of asymmetric equilibria. For example, for \( \varepsilon \) small, \( \alpha_i = 1 - \varepsilon \) and \( \alpha_{-i} = \varepsilon \) comprises a subgame perfect equilibrium where insurer \( -i \) wins both markets and charges \( v \) while insurer \( i \) never enters. Nonetheless, we have:

**Proposition 4.3.** There exists no subgame perfect equilibrium with \( (\alpha_1, \alpha_2) \in (\tilde{\alpha}, 1) \times (\tilde{\alpha}, 1) \).
Proof. See Appendix D.

Insurers will never share the socially efficient amount, $\alpha = 1$, of data with each other unless they are able to sustain collusion (outside of the model). As such, the above results provide an (admittedly stylized) argument for government-mandated datasharing in markets for new or nonstationary risks. Such an approach goes well beyond contributory risk databases (e.g. the Verisk Cyber Data Exchange for anonymized data) in existence today that rely on voluntary contributions, much like the model.\footnote{It is, however, separate from data exchanges for individual, non-anonymized loss histories (e.g. the Comprehensive Loss Underwriting Exchange), which function more like credit scores to combat adverse selection.} A practical implementation of such a scheme would need to respect regulation compelling insurers to describe to their customers exactly how and with whom their data is being shared.

Another implication of the model concerns the McCarran-Ferguson Act and antitrust regulation (e.g. state-level) of insurance more broadly. While in principle, the McCarran-Ferguson Act exempts insurers from federal antitrust regulation, in practice its exemptions apply only to activity that transfers or spreads policy risk, bears on the insurer-insured relationship, and involves multiple insurers; it does not immunize mergers from antitrust actions. Because a merger achieves the socially efficient outcome while datasharing and profitsharing (even if protected by the Act) does not, my results suggest benefits to broadening exemptions in new and nonstationary markets rather than curtailment of the act.\footnote{The Competitive Health Insurance Reform Act of 2020, with stated purpose “to restore the application of the Federal antitrust laws to the business of health insurance to protect competition and consumers” featured a repeal of the McCarran-Ferguson Act for health and dental insurers. Encouragingly, collaboration between firms “collect[ing], compil[ing], or disseminat[ing] historical loss data” was carved out of the repeal, but as I argue, this may still fall sort of achieving efficiency.}

5 Conclusion

In this paper, motivated by the small size of insurance markets for significant but new or nonstationary risks, I model the effects of data on insurance provision. I first develop a general model of decisionmaking (CUE) where model uncertainty, stemming from data paucity, gets penalized; unlike (known) variance, which can be diversified across claims, model uncertainty cannot. I show how CUE, even without classical frictions like moral hazard and adverse selection, can lead an insurance market to fail completely. This creates a data trap – a vicious cycle in which insurers fail to perform the exploratory underwriting that would reduce their model uncertainty and allow them to underwrite in the future. I specialize this model to one of algorithmic decisionmaking, the conservative confidence bound (CCB)
model, and derive the optimal insurance contracts (if any) a monopolist offers: these always feature limits, sometimes feature deductibles, and never feature coinsurance, at once matching several salient features of existing cyberinsurance contracts and generating testable predictions. In particular, I do not invoke the usual frictions of moral hazard and adverse selection—anecdotally of little relevance in these markets—to generate these contract features. Finally, I argue that government data acquisition/generation, mandated data sharing, and broadened antitrust exemptions may be useful tools in expanding such insurance (and lending) markets and producing efficient outcomes.

This paper leaves several fruitful directions for future research. First, and most importantly, predictions of the CCB model vis-a-vis the evolution of loss limits can be tested using longitudinal policy and loss data from brokers. In tandem with solution of the optimal dynamic contract, this exercise would concretely identify CUE from moral hazard and adverse selection as frictions. Second, extending the CUE/CCB frameworks to accommodate robust inference from events (e.g., an agent’s rejection of a contract) may inform a broader agenda interpolating between hyper-strategic and purely algorithmic agent behaviors; I have preliminary work leveraging the tools of differential privacy to define updated distributional envelopes. Finally, a more detailed and dynamic model of datasharing might interact CUE with adverse selection and produce new predictions; these can be compared against contributory risk databases.
References


### A Proofs from Section 2

#### A.1 Proof of Theorem 2.1

I first prove a result on market non-existence when $T = 1$.

**Lemma A.1.** Let $T = 1$ and let $\sup_{\theta \in \Theta} \text{Var}(F^\theta) = \overline{S}$. For any fixed cost $c > 0$ of underwriting, define $\gamma(c)$ such that $\gamma(c)\overline{S} = c$. There exists some $\gamma > \gamma(c)$ such that the insurer with conservative uncertainty evaluation optimally offers no insurance.

**Proof.** First, I argue that $v^*(\gamma) = \sup\{V(-c + \Psi_B - B(X), D) : B(\cdot) \in \mathcal{B}\}$ is continuous in $\gamma$. To do so, observe first that $V$ is continuous in $(B, \gamma)$ as a composition of continuous function; we invoke property (B4). Next, the correspondence $\gamma \mapsto \mathcal{B}$ is constant and hence continuous. Finally, $\mathcal{B}$ is compact (where we use the metric topology induced by the sup norm): by Arzelà-Ascoli it is precompact in $\mathcal{C}(\mathcal{X})$ and it is closed because it consists of $1$-Lipschitz functions. Thus Berge’s theorem applies and $v^*(\gamma)$ is continuous as claimed.

Now observe that for any contract $B$, at $\gamma = \gamma(c)$

$$V(-c + \Psi_B - B(X), D)|_{\gamma = \gamma(c)} = V(-c + E(B^*(X)|D) + \gamma(\text{Var}(X|D) - \text{Var}(X - B(X)|D) - B(X), D).$$

If $B$ is nonconstant, we thus have

$$V(-c + \Psi_B - B(X), D)|_{\gamma = \gamma(c)} < -c + \gamma(c)\text{Var}(X|D) < 0,$$

by conservatism. If $B$ is constant, because $\text{Var}(X - B(X)|D) = \text{Var}(X|D)$, we have

$$V(-c + \Psi_B - B(X), D)|_{\gamma = \gamma(c)} \leq -c < 0,$$

by conservatism. Therefore $v^*(0) < 0$ regardless of the contract $B$. By continuity, there exists some positive $\gamma$ such that $v^*(\gamma) < 0$ which proves the claim. \hfill $\Box$

To tackle the dynamic case $T > 1$, first I introduce a lemma:

**Lemma A.2** (Bellman Optimality Principle). *Suppose, given $D_1$, that $\{B^*_t(\cdot, D_t)\}_{t=1}^T$ is a solution to $\max_{\{B_t(\cdot, D_t)\}_{t=1}^T} W_1$. For all $1 \leq j \leq T$, given $D_j$, its restrictions $\{B^*_t(\cdot, D_t)\}_{t=j}^T$ solve the “continuation” optimizations $\max_{\{B_t(\cdot, D_t)\}_{t=j}^T} W_j$.\textsuperscript{34}*

**Proof.** By induction it suffices to consider the case $T = 2$, and because $j = 1$ is tautological, it suffices to verify the claim for $j = 2$. Let us prove the contrapositive. Suppose that

\textsuperscript{34}Here, I implicitly allow the null contract as a possibility.
\( B^*_{\pi}(x, D_2) \) does not maximize \( W_2 \) given \( D_2 = D_1 \cup X_1 \). Recognizing that the maximized \( W_1 = V(-c + \Psi_1 - B_1(X_1) + \beta W_2(X_1), D_1) \), observe that \( W_2(X_1) \) (obtained at \( B^*_{\pi}(x, D_2) \)) can be increased, given \( X_1 \), for any value of \( X_1 \), by assumption. Thus by monotonicity (B4) of \( V(\cdot, D_1) \), \( W_1 \) can be increased which contradicts optimality.

\[ \square \]

**Remark 15.** If an insurer optimally does not underwrite in some period \( t < T \), he never continues underwriting afterwards: the recursive problem that the insurer solves in period \( t + 1 \) is the same (less a “final period” which adds weakly positive optionality), given the informational assumptions.

Theorem 2.1 then follows from an inductive argument:

**Proof of Theorem 2.1.** By backward induction \((t = T, T-1, \ldots, 1)\), the argument of Lemma A.1 (Berge’s theorem and continuity of the objective) establishes that the optimized \( W^*_t \) is continuous in \( \gamma \). Note that \( W^*_1|_{\gamma=\gamma(c)} < 0 \) by conservatism and fixed costs, so continuity implies there exists a \( \tilde{\gamma} > \gamma(c) \) such that \( W^*_1|_{\tilde{\gamma}} < 0 \). Thus the insurer cannot underwrite in all periods, as the (continuation) value function is implicitly defined to involving underwriting in every period. Once the insurer stops underwriting she never starts again (next period Bellman equation and symmetry) so reapplying this argument a finite number of times implies that there is no underwriting in the first period and hence in any future period.

\[ \square \]

**A.2 Proof of Theorem 2.2**

**Proof of Theorem 2.2.** Let \( B_t(x) = x \) for all \( x, t \). That is, contemplate an insurer who offers full insurance in every period so that her per period cash profit is

\[ \Pi_t(X_t, D_t) = -c + E(X_t|D_t) + \gamma \text{Var}(X_t|D_t) - X_t. \]

By consistency (B2) and the consistency of posterior probability measures (Bernstein-von Mises), for all \( \varepsilon_1 > 0 \) there exists \( s = |D_1| \) such that for all \( 1 \leq t \leq T \), and hence \( |D_t| \geq s \),

\[ P \left( |V(\Pi_t(X), D_t) - E_{F^{\theta,N}}(\Pi_t(X)|D_t)| > \varepsilon_1 \right) < \frac{\varepsilon}{2T}, \tag{7} \]

\[ P \left( |E(X|D_t) - E_{F^{\theta,N}}(X|D_t)| > \varepsilon_1 \right) = P \left( |E(X|D_t) - E_{F^{\theta,N}}(X)| > \varepsilon_1 \right) < \frac{\varepsilon}{4T}. \tag{8} \]

\[ P \left( |\text{Var}(X|D_t) - \text{Var}_{F^{\theta,N}}(X|D_t)| > \varepsilon_1 \right) = P \left( |\text{Var}(X|D_t) - \text{Var}_{F^{\theta,N}}(X)| > \varepsilon_1 \right) < \frac{\varepsilon}{4T}. \tag{9} \]

The conditioning on \( D_t \) in expectations with respect to \( F^{\theta,N} \) indicates that only randomness in \( X \) is being integrated over, and no updating is occurring. In other expectations the data
is being used to update $\pi$. Pick $\varepsilon_1 < \text{Var}_{F^\theta N}(Y)$. Note (writing $Y$ generated independently according to the same process as $\{X_t\}$ for expository clarity)

$$E_{F^\theta N}(\Pi_t(X)|D_t)) = E_{F^\theta N}(-c + E(Y|D_t) + \gamma \text{Var}(Y|D_t) - X|D_t)$$

$$= -c + \gamma \text{Var}(Y|D_t) + E(Y|D_t) - E_{F^\theta N}(X)$$

$$\geq -c + \gamma \text{Var}_{F^\theta N}(Y) + \gamma (\text{Var}(Y|D_t) - \text{Var}_{F^\theta N}(Y)) + (E(Y|D_t) - E_{F^\theta N}(X))$$

Then, with probability exceeding $1 - \varepsilon/2T$ (using a union bound),

$$E_{F^\theta N}(\Pi_t(X)) \geq -c + \gamma \text{Var}_{F^\theta N}(Y) - (1 + \gamma)\varepsilon_1 > \varepsilon_1$$

so long as

$$\gamma > \frac{c + 2\varepsilon_1}{\text{Var}_{F^\theta N}(Y) - \varepsilon_1}.$$

A union bound using Equation (7) implies

$$P(V(\Pi_t(X), D_t) > 0) > 1 - \frac{\varepsilon}{T}.$$ 

Applying another union bound,

$$P\left(\{V(\Pi_t(X), D_t) > 0 \text{ for all } t\}\right) > 1 - \varepsilon.$$

Thus, with high probability, writing full insurance in each period will earn the insurer positive cash profits in each period. Suppose an insurer (optimally) does not underwrite in every period; say that period $m < T$ is the last period in which she underwrites (per earlier arguments, once the insurer stops underwriting, she never underwrites again). Consider her continuation problem starting at time $t = m + 1$ : offering full insurance generates profit, which contradicts optimality. Thus, with probability exceeding $1 - \varepsilon$, the insurer underwrites in each period, as desired. \hfill \square

**B Discussion: Competition in the Context of Section 2**

In what follows, I sketch the basic elements required to extend the general model of Section 2.2.1 to a setting with competition, and the corresponding results I expect. Doing so facilitates an analysis of datasharing through its interaction with distributional uncertainty. Standard models of datasharing in insurance and lending ignore this channel, focusing in-
Consider two competing insurers $A$ and $B$ with different “starter” data sets $D^A$ and $D^B$ and a consumer with the union of the datasets. These datasets are private information, but their sizes $m^A, m^B > 0$ are known. For now, let $m^A = m^B = m$ for simplicity. Supposing, loosely, that competition only worsens an insurer’s payoff, consider four cases of interest (where Case 2 may subsume Case 1; the rest are logically disjoint):

1. Neither insurer $i \in \{A, B\}$ would write if a monopolist assuming both she and the customer possesses data $D^i$.

2. Neither insurer $i \in \{A, B\}$ would write if a monopolist, assuming that the customer possesses data $D^A \cup D^B$.

3. One insurer $i \in \{A, B\}$ would write if a monopolist, assuming that the customer possesses data $D^A \cup D^B$.

4. Both insurers $i \in \{A, B\}$ would write if monopolists, assuming that the customer possesses data $D^A \cup D^B$.

By Theorem 2.1, Case 1 is a possibility. Loosely, a configuration (of preference parameters, uncertainty, etc.) in Case 1 should reasonably belong to Case 2 as well: Case 2 simply introduces adverse selection. A formal analysis of these cases requires:

1. A notion of valuation based on multiple sources of uncertainty

2. A notion of “updating” given the information that the buyer accepts the offer – i.e. of appending that information to the dataset and correspondingly extending $V$’s definition to include this different sort of information

Continuing the informal argument, if a configuration belongs to Case 2, the problem worsens in competition. Thus an analog of Theorem 2.1 about the possibility of uncertainty-induced market failure is true in markets with competition. An analog of Theorem 2.2 should also hold for the model with competition, where this data could be provided by the government.

If, further, it is the case that for some configuration in Case 2 that a hypothetical monopolist insurer possessing $D^A \cup D^B$ would write if the customer possesses data $D^A \cup D^B$.

\textsuperscript{35}Income sharing agreements and thin file credit are relevant examples here.
then so would competing insurers each possessing $D^A \cup D^B$, though each would achieve a zero valuation. This means that a \textit{data sharing policy} could ameliorate a situation of market failure, at least for one period. Stronger assumptions are necessary to make this amelioration perpetual. On the other hand, configurations corresponding to Case 3 where a single insurer would underwrite under competition, incurring "cash" losses in the first several periods imply situations where a \textit{future data sharing policy} would eliminate any market, present or future by eliminating the value of the insurer’s exploratory underwriting investment.

\section{Proofs from Section 3}

\subsection{Construction of Optimal Contract}

In this section, I construct the solution to the CCB monopolist insurer’s problem. This construction will immediately imply \textbf{Theorem 3.1}.

To start, consider the relaxed problem from \textbf{Section 3.2.1} which removes the $(\nu)$ constraints. Denoting the multiplier associated with the relevant $\mu$ constraint by $\mu_i$ and likewise for $\lambda$ constraints (with $i = 0$ corresponding to $b_1 \geq 0$) we may form the Lagrangian

$$L = \left( -\epsilon_n b_n + \frac{1}{n} \sum_{j=1}^{k} b_j + \left( \epsilon_n - \frac{k}{n} \right) b_k + \frac{\gamma}{n} \left( \text{Var}(X) - \sum_{i=1}^{n} (x_i - \bar{x})^2 \right) - \left( \sum_{j=1}^{n-1} \lambda_j (b_{j+1} - b_j) + \lambda_0 (b_1 - 0) + \sum_{j=1}^{n} \mu_j (x_j - b_j) \right) \right).$$

(10)

Differentiating with respect to $b_i$ and simplifying gives the following first order conditions
\[(i = 1) \quad \frac{1}{n} + \frac{\gamma}{n} \left(2(x_1 - b_1 - (\bar{x} - \bar{b}))\right) + \lambda_1 - \lambda_0 + \mu_1 = 0\]  
\[(2 \leq i \leq k) \quad \frac{1}{n} + \frac{\gamma}{n} \left(2(x_i - b_i - (\bar{x} - \bar{b}))\right) - \lambda_{i-1} + \lambda_i + \mu_i = 0\]  
\[(i = k + 1) \quad \left(\varepsilon - \frac{k}{n}\right) + \frac{\gamma}{n} \left(2(x_{k+1} - b_{k+1} - (\bar{x} - \bar{b}))\right) - \lambda_k + \lambda_{k+1} + \mu_{k+1} = 0\]  
\[(k + 1 < i \leq n - 1) \quad \frac{\gamma}{n} \left(2(x_i - b_i - (\bar{x} - \bar{b}))\right) - \lambda_{i-1} + \lambda_i + \mu_i = 0\]  
\[(i = n) \quad \varepsilon - \frac{\gamma}{n} \left(2(x_n - b_n - (\bar{x} - \bar{b}))\right) - \lambda_{n-1} + \mu_n = 0\]

Ignoring all of the constraints (i.e. assuming they are slack and that the associated Lagrange multipliers are zero):

\[(1 \leq i \leq k) \quad x_i - b_i - (\bar{x} - \bar{b}) = \frac{-1}{2\gamma}\]
\[(i = k + 1) \quad x_{k+1} - b_{k+1} - (\bar{x} - \bar{b}) = \left(\frac{k}{n} - \varepsilon\right) \frac{n}{2\gamma}\]
\[(k + 2 \leq i \leq n - 1) \quad x_i - b_i - (\bar{x} - \bar{b}) = 0\]
\[(i = n) \quad x_n - b_n - (\bar{x} - \bar{b}) = \frac{n\varepsilon}{2\gamma}\]

Difference consecutive equations to see immediately that \(b_i = b(x_i)\) has slope 1 for the two “constant” ranges of values. At \(i = k + 1\) (that is, \(k \to k + 1\) and \(k + 1 \to k + 2\)) and \(i = n\) (that is, \(n - 1 \to n\)) we see that \(x_i - b_i\) jumps (relative to its average) meaning that the contract grows at slope lesser than 1 at those points. Our relaxed solution thus satisfies all \((\nu)\) constraints and trivially satisfies all \((\mu)\) constraints. This leaves only \(b_1 = b(x_1)\) as a free parameter, and note that any \(b_1 \in [0, x_1]\) yields the same value of the objective function without violating these constraints: changing \(b_1\) does not affect the expected net profit to the insurer because what is gained in the “early region” is lost by the parallel shift up in costs incurred at \(X(n)\) and the risk premium term is unchanged as we are simply adding a constant inside our variance calculation.
The final family of constraints are the $(\lambda)$ constraints, which again do not depend on $b_1$. Rearranging the differenced equations above, we have

\[(1 \leq j \leq k) \quad b_j = b_1 + (x_j - x_1) \quad (16)\]
\[(j = k + 1) \quad b_{k+1} = b_1 + (x_{k+1} - x_1) + \left(-\frac{1}{2\gamma} - \left(\frac{k}{n} - \varepsilon_n\right)\frac{n}{2\gamma}\right) \quad (17)\]
\[(j = k + 2) \quad b_{k+2} = b_1 + (x_{k+2} - x_1) - \frac{1}{2\gamma} \quad (18)\]
\[(k + 3 \leq j \leq n - 1) \quad b_j = b_1 + (x_j - x_1) - \frac{1}{2\gamma} \quad (19)\]
\[(j = n) \quad b_n = b_1 + (x_n - x_1) - \frac{1}{2\gamma} - \frac{n\varepsilon_n}{2\gamma} \quad (20)\]

In particular, we must check at the kinks that monotonicity is not violated. That is,

\[b_{k+1} \geq b_k \iff x_{k+1} - x_k \geq \frac{1}{2\gamma} + \left(\frac{k}{n} - \varepsilon_n\right)\frac{n}{2\gamma} \quad (21)\]
\[b_{k+2} \geq b_{k+1} \iff x_{k+2} - x_{k+1} \geq \left(\varepsilon_n - \frac{k}{n}\right)\frac{n}{2\gamma} \quad (22)\]
\[b_n \geq b_{n-1} \iff x_n - x_{n-1} \geq \frac{n\varepsilon_n}{2\gamma} \quad (23)\]

The unconstrained optimization implies

**Lemma C.1.** If Equation (21)-Equation (23) hold, the equations Equation (16)-Equation (20) define an optimal contract with arbitrary $b_1 \in [0, x_1]$.

**Remark 16.** The result of Lemma C.1 would be an SFSF contract in the language of Theorem 3.1.

What if some of Equation (21)-Equation (23) fail? That is, what if the unconstrained solution violates monotonicity at one of these three segments?

I consider violations of Equation (23) separately from Equation (21) and Equation (22).

1. First, I solve the appropriate contract flattening (backward from the right) for when Equation (23) is violated but Equation (21) and Equation (22) are not; this flattened region may (A) not overlap with $[0, x_{k+1}]$, (B) may overlap. In all cases I show how to proceed, and in particular whether Equation (22) or Equation (21) is violated is immaterial to the analysis of (B).
2. Then I consider the analysis of violations of Equation (22) or Equation (21).

- If the appropriate flattening around this region does not interact with the possibly flattened region corresponding to Equation (23) we are done.
- If the flattening of this region does interact with the possibly flattened region corresponding to Equation (23), I return to the analysis of case (B) and use the solution there, if it exists.

3. Finally, in case (B) it is possible we wind up with no solution, corresponding to a noncontracting outcome.

In all cases, the “flattenings” imply nonzero values only for Lagrange multipliers $\lambda_i$ and I correspondingly produce a stationary point $(b^*, \lambda^*)$ of the Lagrangian. I conclude using the KKT theorem noting the quadratic objective and linear constraints. The argument given above, coupled with the finding of the claimed stationary points, completes the proof of Theorem 3.1.

**Violations of Equation (23):** Let

$$s = \max \left\{ t : \sum_{j=t+1}^{n} (x_j - x_t) \geq \frac{n\varepsilon_n}{2\gamma} \right\}$$

Thus $s < n - 1$ if and only if Equation (23) is violated. Given $s$ as above, define

$$\Delta = \frac{1}{n-s} \left( \sum_{j=s+1}^{n} (x_j - x_s) - \frac{n\varepsilon_n}{2\gamma} \right) \geq 0$$

whereby

$$\sum_{j=s+1}^{n} (x_j - x_s - \Delta) = \sum_{j=s+1}^{n} (x_j - x_s) - (n-s)\Delta = \frac{n\varepsilon_n}{2\gamma}$$  

(24)

Readily by the definition of $s$,

$$\sum_{j=s+2}^{n} (x_j - x_{s+1}) < \frac{n\varepsilon_n}{2\gamma} = \sum_{j=s+1}^{n} (x_j - x_s) - (n-s)\Delta$$

$$\implies (n-s)(x_{s+1} - x_s) > (n-s)\Delta$$

$$\implies x_{s+1} - x_s - \Delta > 0$$
where the equality follows from substituting Equation (24). The index \( s \) (or some auxiliary index \( 0 \leq \nu \leq s \)) will help us define a flat region starting at \( b_{\nu+1} \), ultimately implying that the policy has a limit there. When \( \nu = 0 \) this implies no contract will obtain (the corresponding family of multipliers \( \{\lambda\} \) will be computable).

**Case A:** \( s > k + 1 = [n\varepsilon_n] + 1 \) (see illustration in Figure 6):

Subcase AA: Equation (21) and Equation (22) hold: Consider Equation (10) and relax it further, keeping only constraints in Equation (4) corresponding to the Lagrange multipliers \( \{\lambda_j\}_{j=s+1}^{n-1} \) and zeroing out the other multipliers in the Lagrangian. This produces a system analogous to Equation (11)-Equation (15) which I reproduce here for convenience, with the reminder \( \lambda_s = 0 \):

\[
\begin{align*}
(1 \leq i \leq k) & \quad \frac{1}{n} + \frac{\gamma}{n} \left( 2(x_i - b_i - (\bar{x} - \bar{b})) \right) = 0 \quad (25) \\
(i = k + 1) & \quad \left( \varepsilon_n - \frac{k}{n} \right) + \frac{\gamma}{n} \left( 2(x_{k+1} - b_{k+1} - (\bar{x} - \bar{b})) \right) = 0 \quad (26) \\
(k + 1 < i \leq s) & \quad \frac{\gamma}{n} \left( 2(x_i - b_i - (\bar{x} - \bar{b})) \right) = 0 \quad (27) \\
(s + 1 \leq i \leq n - 1) & \quad \frac{\gamma}{n} \left( 2(x_i - b_i - (\bar{x} - \bar{b})) \right) - \lambda_{i-1} + \lambda_i = 0 \quad (28) \\
(i = n) & \quad \varepsilon_n + \frac{\gamma}{n} \left( 2(x_n - b_n - (\bar{x} - \bar{b})) \right) -\lambda_{n-1} = 0 \quad (29)
\end{align*}
\]

Let \( \lambda_i \) be valued according to the recursive relations below (and zero otherwise):

\[
\begin{align*}
(i = n - 1) & \quad \lambda_{n-1} = -\varepsilon_n + \frac{2\gamma}{n} (x_n - x_s - \Delta) \\
(s \leq i \leq n - 2) & \quad \lambda_i = \lambda_{i+1} + \frac{2\gamma}{n} (x_{i+1} - x_s - \Delta)
\end{align*}
\]

Note, as a check, that plugging in \( i = s \) yields zero, so indeed the constraint \( b_s \leq b_{s+1} \) is slack. Note also that all \( \lambda_i \leq 0 \) because of the signing of my Lagrangian. Further, let \( \{b_i\}_{i=1}^{n} \) be valued according to the following equations:
\[(1 \leq j \leq k) \quad b_j = b_1 + (x_j - x_1) \quad (30)\]
\[(j = k + 1) \quad b_{k+1} = b_1 + (x_{k+1} - x_1) + \left( -\frac{1}{2\gamma} - \left( \frac{k}{n} - \varepsilon_n \right) \frac{n}{2\gamma} \right) \quad (31)\]
\[(k + 2 \leq j \leq s) \quad b_j = b_1 + (x_j - x_1) - \frac{1}{2\gamma} \quad (32)\]
\[(s + 1 \leq j \leq n) \quad b_j = b_1 + (x_{s+1} - x_1) - \frac{1}{2\gamma} - \Delta \quad (33)\]

It is routine to check that \(\{\lambda_i\}, \{b_i\}\) given above constitute a solution to the system of Equation (25)-Equation (29) which, by the KKT theorem, are necessary conditions for a solution to the insurer’s optimization problem. In fact, these are also sufficient conditions. So for all choices \(0 \leq b_1 \leq x_1\) the contract given is an optimal contract, which the insurer will underwrite if the value of the objective function exceeds the fixed cost \(c\). Therefore, we have

**Lemma C.2.** Suppose Equation (23) is violated but neither Equation (21) nor Equation (22) is. Suppose \(s > k + 1\). Then Equation (30)-Equation (33) solve the CCB insurer’s problem, with arbitrary choice \(b_1 \in [0, x_1]\). There exists a SFSF solution with \(b_1 = x_1\).

Qualitatively, observe that this policy has a loss limit at \(b_{s+1} = b_1 + (x_{s+1} - x_1) - \frac{1}{2\gamma} - \Delta\). Elsewhere, the slope is 1 (and hence the contract is linear by constraint) everywhere except on the segments \(x_k \to x_{k+1}, x_{k+1} \to x_{k+2}\) and \(x_s \to x_{s+1}\) where the slope is less than one and thus the shape of the interpolating function is undetermined (though economically irrelevant). Finally, note that there is no unique optimal contract, as choice of \(b_1\) is arbitrary.

**Subcase AB:** Equation (21) or Equation (22) is violated: Proceed analogously to Subcase AA for determining values \(b_n, b_{s+1}, \ldots, b_{n}\). See the comments on Violations of Equation (21) or Equation (22) for details on constructing the rest of contract and incorporating the extra binding monotonicity constraints in the “earlier” part of the contract; Figure 7 provides an example. As noted earlier, this may default analysis into Case B, depending on compatibility of flattened regions. I eschew a formal lemma here.

**Case B:** \(1 \leq s \leq k + 1\) or \(s\) does not exist\(^{36}\). Compute

\(^{36}\)That is, \(\sum_{j=2}^n (x_j - x_1) < n\varepsilon_n/2\gamma\). Evidently \(\tilde{s}\) and \(\bar{s}\) cannot exist. It is clear that all monotonicity constraints bind and the insurer picks a completely flat contract \(b(x) = \bar{b} \leq x_1\), with the obvious corresponding choice of multipliers \(\{\lambda_i\}\). The same is true when \(\tilde{s}\) (resp. \(\bar{s}\)) do not exist even when \(s\) (resp. \(s, \tilde{s}\)) do.
Figure 6: Imposing monotonicity given violation of Equation (23): The image on the left shows the contract (in the transformed $x_i - b_i - (\bar{x} - \bar{b})$ space) that solves the unconstrained problem, including infeasible segments where the slope increases faster than one. This includes the segments $\{(x_k, x_{k+1}), (x_{k+1}, x_{k+2}), (x_{n-1}, x_n)\}$ so we are in Case AB. The red dots show how, from the right, the final region of the contract is amended to a feasible one in a way that preserves $\bar{x} - \bar{b}$; this corresponds to flattening the contract such as to impose a loss limit. The text provides details on reading Lagrange multipliers off this diagram.

$$
\tilde{s} = \max \left\{ t : \sum_{j=t+1}^{n} \left( x_j - x_t - \left( \varepsilon_n - \frac{k}{n} \right) \left( \frac{n}{2\gamma} \right) \right) / 2 \gamma \geq \frac{k}{2\gamma} \right\},
$$

$$
\tilde{\Delta} = \frac{1}{n - \tilde{s}} \left( \sum_{j=\tilde{s}+1}^{n} \left( x_j - x_{\tilde{s}} - \left( \varepsilon_n - \frac{k}{n} \right) \left( \frac{n}{2\gamma} \right) \right) - \frac{k}{2\gamma} \right) \geq 0
$$

if $\tilde{s}$ exists (if not, the optimal contract is flat). Evidently $\tilde{s} \leq s$ because there is at least one summand. If still $\tilde{s} = k + 1$, proceed as above in Subcase AA: the segment $k + 1 \rightarrow k + 2$ gets flattened (more) and the policy limit is $b(x_{k+2})$: the contract is flat thereafter.\(^{37}\) If $\tilde{s} < k + 1$, compute

$$
\tilde{\tilde{s}} = \max \left\{ t : \sum_{j=t+1}^{n} \left( x_j - x_t - \frac{1}{2\gamma} \right) \geq \frac{t}{2\gamma} \right\}, \quad \tilde{\Delta} = \frac{1}{n - \tilde{\tilde{s}}} \left( \sum_{j=\tilde{\tilde{s}}+1}^{n} \left( x_j - x_{\tilde{\tilde{s}}} - \frac{1}{2\gamma} \right) - \frac{\tilde{\tilde{s}}}{2\gamma} \right) \geq 0
$$

if $\tilde{\tilde{s}}$ exists (if not, the optimal contract is flat). Again, $\tilde{\tilde{s}} \leq \tilde{s}$: move over the subtracted terms in each summand to the other side to see that the condition gets harder to satisfy for any

\(^{37}\)I leave the computation of the exact Lagrange multipliers for verification of optimality to the reader.
If $\tilde{s} \geq 1$ exists, construct the contract exactly as in Subcase A: the contract at all points greater than or equal to $x_{\tilde{s}+1}$ is flat and the contract at all points lesser has slope 1, except for the segment connecting $x_{\tilde{s}}$ to $x_{\tilde{s}+1}$ which has slope weakly greater than 0 but less than 1. Specifically, let $\lambda_i$ be valued according to the recursive relations below (and zero otherwise):

\[
(i = n - 1) \quad \lambda_{n-1} = \frac{2\gamma}{n} \left( -\tilde{\Delta} + (x_n - x_{\tilde{s}}) - \frac{n\varepsilon_n}{2\gamma} - \frac{1}{2\gamma} \right)
\]

\[(k + 1 \leq i \leq n - 2) \quad \lambda_i = \lambda_{i+1} + \frac{2\gamma}{n} \left( -\tilde{\Delta} + (x_{i+1} - x_{\tilde{s}}) - \frac{1}{2\gamma} \right)
\]

\[(i = k) \quad \lambda_k = \lambda_{k+1} + \frac{2\gamma}{n} \left( -\tilde{\Delta} + (x_{k+1} - x_{\tilde{s}}) - \frac{1}{2\gamma} - \left( \frac{k}{n} - \varepsilon_n \right) \frac{n}{2\gamma} \right)
\]

\[(s \leq i \leq k - 1) \quad \lambda_i = \lambda_{i+1} + \frac{2\gamma}{n} \left( -\tilde{\Delta} + (x_{i+1} - x_{\tilde{s}}) - \frac{1}{2\gamma} + \frac{1}{2\gamma} \right)
\]

As a check, plugging in $i = s$ yields zero; the constraint $b_s \leq b_{s+1}$ is slack. As earlier, all $\lambda_i \leq 0$ because of the signing of my Lagrangian. To be explicit, let $\{b_i\}_{i=1}^n$ be valued according to the following equations, for $b_1 \in [0, x_1]$.

\[(1 \leq j \leq \tilde{s}) \quad b_j = b_1 + (x_j - x_1) \quad \text{(34)}
\]

\[j = \tilde{s} + 1 \quad b_{\tilde{s}+1} = b_{\tilde{s}} + (x_{\tilde{s}+1} - x_{\tilde{s}}) - \tilde{\Delta} \quad \text{(35)}
\]

\[(\tilde{s} + 1 < j \leq n) \quad b_j = b_{\tilde{s}+1} \quad \text{(36)}
\]

Whether or not Equation (21) and Equation (22) are violated is irrelevant. So in a manner analogous to earlier, we may pick $b_1 = x_1$ and get

Lemma C.3. Suppose Equation (23) is violated.

1. If $1 \leq s \leq k + 1$, $\tilde{s} < k + 1$, and $\tilde{s}$ exists, the solution to the insurer’s problem is given by Equation (34)–Equation (36), whether or not Equation (21) and Equation (22) are violated. Picking $b_1 = x_1$, there exists an optimal contract with configuration SF.

2. If $s$ does not exist, or if $s$ exists and $\tilde{s}$ does not exist, or if $(s, \tilde{s})$ exist and $\tilde{s}$ does not exist, then the optimal contract is the zero contract (configuration F).

Remark 17. The statement of this lemma omits the case $\tilde{s} = k + 1$ because I do not explicitly provide equations for it. The result is exactly analogous, and a SF contract obtains.
**Violations of Equation (21) and Equation (22)** There are three possibilities:

1. Equation (23) is not violated

2. Equation (23) is violated: we are in Subcase AB

3. Equation (23) is violated: we are in Case B.

Recall that if dealing with a violation of Equation (23) and in Case B, we simply proceed as indicated there. So we restrict attention to the first two cases, noting that solving Subcase AB may require a solution per Case B (computing \( \tilde{s} \) and proceeding). The goal will be to flatten the contract to rectify nonmonotonicities in the relaxed solution’s \( b_k \to b_{k+1} \) and \( b_{k+1} \to b_{k+2} \) (corresponding to imposing binding monotonicity constraints associated with Equation (21) and Equation (22)) and simultaneously preserve the other (flattened) first order conditions (in a sense to be made precise).

Define

\[
h(x) = \begin{cases} 
  -\frac{1}{2^\gamma} & \text{if } x = x_i, \quad 1 \leq i \leq k \\
  \left( \frac{k}{n} - \varepsilon_n \right) \frac{n}{2^\gamma} & \text{if } x = x_{k+1} \\
  0 & \text{if } x = x_i, \quad i > k + 1
\end{cases}
\]

to assume the value of \( x_i - b_i - (\bar{x} - \bar{b}) \) that the completely unconstrained contract\(^{38}\) does at each \( x_i \). For \( \beta \in [x_{k+1}, x_{k+1} + \frac{1}{2^\gamma}] \), consider the parallelogram in this space determined by the lines \( y = x - \beta \) as pictured in Figure 7. Picking a particular \( \beta \) defines a flattening of the contract (taking evaluations along the line) over the relevant domain of \( x_i \) (so monotonicity constraints are imposed and the contract shape is feasible). Doing so will not, in general, preserve \( \bar{x} - \bar{b} \) over the amended region so optimality, as per KKT, need not obtain.

Thus motivated, define\(^{39}\)

\[
\mathcal{I}(\beta) = \int_{\beta - \frac{1}{2^\gamma}}^{\beta} (x - \beta - h(x))dF_n(x),
\]

where \( F_n \) is the empirical CDF. Note \( \mathcal{I}(\beta) \) is continuous for \( \beta \in [x_{k+1}, x_{k+1} + \frac{1}{2^\gamma}] \). Further, \( \mathcal{I}(x_{k+1}) > 0 \) and \( \mathcal{I} \left( x_{k+1} + \frac{1}{2^\gamma} \right) < 0 \) so by the intermediate value theorem, there exists some \( \beta^* \in (x_{k+1}, x_{k+1} + 1/2^\gamma) \) such that \( \mathcal{I}(\beta^*) = 0. \)

---

\(^{38}\)In particular, ignore any monotonicity adjustment made in the previous section.

\(^{39}\)For the sake of simplicity, I assume \( x_{k+1} + 1/2^\gamma < x_n \).
Figure 7: Imposing monotonicity given violations of Equation (21) and Equation (22). These diagrams complete the derivation of the optimal contract in Figure 6. The first image shows the set of lines $\beta \in [x_{k+1}, x_{k+1} + \frac{1}{2\gamma}]$. The second through fourth show cases where $I(\beta) > 0$, $I(\beta) < 0$, and $I(\beta) = 0$ respectively. The fifth image shows the resulting contract in $x_i - b_i - (\overline{x} - \overline{b})$ space, and the final shows the contract in usual coordinates, thus completing the analysis of an example belonging to Subcase AB with SFSF configuration.
Case I: $\beta^* < \frac{1}{2\gamma}$. In the “dot diagram,” this means that at $x_1$, the implied $x_1 - b_1 - (\bar{x} - \bar{b})$ does not paste to $-1/2\gamma$. To satisfy equations Equation (11)-Equation (15) (fixing all $\mu_i = 0$) this simply requires picking the appropriate multiplier $\lambda_1 < 0$ (recall the signing of the Lagrangian) and likewise “flattening forward.” In particular, if the front-flattening includes index $s$, we return to Case B and solve for $\bar{s}$ if possible. If not, we solve the back-flattening separately and paste solutions. The solution thus obtained is either of FSF or SF/F configuration.

Case II: $\beta^* \geq \frac{1}{2\gamma}$. Note that $\beta^*$ defines indices $L < k + 1$ and $R > k + 1$ such that the augmented $x - b - (\bar{x} - \bar{b})$ has slope 1 between $x_{L+1}$ and $x_{R-1}$ and slope greater than zero but weakly less than one between $x_L$ and $x_{L+1}$ and $x_{R-1}$ and $x_R$. Formally,

$$L(\beta^*) = \min \left\{ i : x_i - \beta^* \geq \frac{-1}{2\gamma} \right\} - 1, \quad R(\beta^*) = \max \{ i : x_i - \beta^* \leq 0 \} + 1$$

Subcases depend on $R$ and the specific violation of Equation (23).\footnote{If $R > n$ the optimal contract is the flat contract.} If Equation (23) is violated, define $s$ as earlier and replace by $\bar{s}$ or $\tilde{s}$ where applicable (if these are used in the optimal backflattening).

**Subcase II.1:** If Equation (23) is unviolated, apply the solution above to the relevant region and for higher indices, use the differenced version of the unconstrained solution to determine $b_{i+1} - b_i$ and hence $b_{i+1}$. This produces a SFSF contract.

**Subcase II.2:** $R \leq s$: this is the “compatible case.” Apply the solution as above and determine higher terms $b_{R+1}$ onwards by the differenced form of Equation (30)-Equation (33). This produces a SFSF contract.

**Subcase II.3:** $R > s + 1$: proceed to Case B and compute $\tilde{s}$. This produces a contract with configuration F or SF.

Putting this all together,

**Lemma C.4.** Suppose Equation (21) or Equation (22) is violated. The insurer’s problem is solved either by:

- Using the front-flattening procedure and appending the differenced unconstrained results (if Equation (23) unviolated)
• Using the front-flattening procedure and appending the constrained (backflattened) results from above (if Equation (23) violated and flattened regions are compatible).

• Reverting to Case B, computing \( \tilde{s} \), and deriving the corresponding globally optimal contract (if Equation (23) violated and flattened regions are incompatible).

The contract produced can have configurations F, SF, FSF, or SFSF.

These results imply the stated Theorem 3.1: the existence of a limit was clear prima facie (Remark 12). All possible contract configurations involve and end in a flat region, implying a policy limit \( L < x_n \). While this is obvious from the first order conditions, the rest of the theorem is proved by collecting the mutually exhaustive cases above.

C.2 Asymptotic Impairment of Coverage

In this subsection, I present a partial characterization of how asymptotic coverage \( \lim_{n \to \infty} L^*/x_n \) depends upon tail behavior of \( F \). Specifically, I present results for \( F \) in the Fréchet maximum domain of attraction for \( \alpha > 2 \) (light-tailed) for \( \alpha < 1 \) (heavy-tailed). For the Gumbel MDA (very light-tailed), it is possible to strengthen Lemma C.5 but a different approach is needed for a formal characterization. See Embrechts et al. (2013) for definitions.

Here is a lemma ruling out locations for \( L^* \) in the light-tailed case:

**Lemma C.5.** Fix \( 0 < \varepsilon, l \in \mathbb{Z}^+ \). Let \( F \) be such that for \( X, X_1, \ldots, X_n \sim iid F, E(X^{2+\varepsilon}) < \infty \). Then, as \( n \to \infty \)

\[
P \left( L^* \in (X_{(n-l)}, X_{(n)}) \right) \to 0.
\]

**Proof.** First, for \( Z_n = \max(X_1, \ldots, X_n) \), note that that \( E(Z_n) \leq O(n^{1/(2+\varepsilon)}) \) by concavity of \( x \mapsto x^{1/(2+\varepsilon)} \) and bounding \( Z_n^{2+\varepsilon} \leq \sum_{i=1}^n X_i^{2+\varepsilon} \). Then Markov’s inequality gives

\[
P \left( Z_n > \log(n)O(n^{1/(2+\varepsilon)}) \right) \leq O \left( \frac{1}{\log(n)} \right).
\]

A necessary condition for \( L \in (X_{(n-l)}, X_{(n-l+1)}) \) to be a solution is that \( s = n - l \), i.e. the sum condition is satisfied there. Rearranging,

\[
P \left( L^* \in (X_{(n-l)}, X_{(n-l+1)}) \right) \leq P \left( \sum_{j=n-l+1}^n (X_{(j)} - X_{(n-l)}) \geq \frac{n \varepsilon n}{2\gamma} \right) \to 0,
\]

as

\[
\frac{c}{\sqrt{n}} \sum_{j=n-l+1}^n (X_{(j)} - X_{(n-l)}) \leq \frac{c}{\sqrt{n}} lZ_n = o_p(1)
\]

55
The claim follows from taking a union bound.

This result implies:

**Proposition C.1.** For $F$ in the Frechet MDA with $\alpha > 2$,

$$\liminf_{n \to \infty} P(X_{(n)}/L^* > z) \geq 1 - (1 - z^{-\alpha})^l \text{ for } z > 1.$$

**Proof.** If $F$ in the maximum domain of attraction of a Frechet distribution with $\alpha > 2$, then it satisfies the hypotheses of Lemma C.5. Therefore for any fixed $l$, we have $P(L^* \in (X_{(n-l)}, X_{(n)}) \to 0$. So,

$$P \left( \frac{X_{(n)}}{L^*} > \frac{X_{(n)}}{X_{(n-l)}} \right) \to 1.$$

By Theorem 3.1(2) of Balakrishnan and Stepanov (2008),

$$P \left( \frac{X_{(n)}}{X_{(n-l)}} > z \right) \to \sum_{i=0}^{l-1} \binom{l}{i} (1 - z^{-\alpha})^i z^{-\alpha(l-i)} = 1 - (1 - z^{-\alpha})^l \text{ for } z > 1.$$

The heavy-tailed case admits a more direct proof.

**Proposition C.2.** For $F$ in the Frechet MDA with $\alpha < 1$,

$$L^*/X_{(n)} \xrightarrow{P} 1.$$

**Proof.** Note that

$$|X_{(n)} - L^*| \leq \frac{Cn}{2\gamma}$$

by definition of $s, \tilde{s}$ and a crude bound on the size of the maximal front-flattening. Per Nagaraja et al. (2015) and Embrechts et al. (2013)(see Corollary 4.2.13),

$$\frac{X_{(n)} - X_{(n-1)}}{b_n} \xrightarrow{d} W_1 - W_2$$

where $b_n = F^{-1} \left( 1 - \frac{1}{n} \right) = n^{1/\alpha}L_1(n)$ for $L_1 \in \mathcal{R}_0$ are the norming constants in the FTG theorem and the random variables $(W_1, W_2)$ have the joint representation

$$(W_1, W_2) \overset{d}{=} (Z_1^{-1/\alpha}, (Z_1 + Z_2)^{-1/\alpha})$$

56
for $Z_i \sim \text{iid} \ Exp(1)$. But $\alpha < 1$, so this scaling implies $L^*/X(n) \xrightarrow{d} 1$ which implies convergence in probability.

\section*{D \ Proofs from Section 4}

\textit{Proof of Proposition 4.1.} Coupling makes it immediate that $\Pi_{m}^{F_n}(D) > \Pi_{m}^{U_n}(D)$ for all $m$. Theorem 2.2 implies the claim: for $m$ sufficiently large, $\Pi_{m}^{F_n}(D) > 0$ and picking

$$\max\left(0, \frac{\Pi_{m}^{U_n}(D)}{m}\right) < \eta_I < \eta_G < \frac{\Pi_{m}^{F_n}(D)}{m}$$

\textit{Proof of Proposition 4.2.} Regardless of choice of $\{\gamma_i\}_{i=1,2}$, subgame perfection means that the usual Bertrand outcome obtains in each market once fractions $\{\alpha_i\}_{i=1,2}$ are picked. If these fractions are less than or equal to $\tilde{\alpha}$, insurer $i$ prices at $\max(c_{i,i}, v)$ in market $i$; insurer $-i$ prices at an irrelevant $p_{-i,i} \geq c_{-i,i} = c(\alpha_i, n) \geq v$ with equality only when $\alpha_i = \tilde{\alpha}$. If $\alpha_i > \tilde{\alpha}$ then $c_{-i,i} < v$ meaning that insurer $-i$ will charge $\max(c_{-i,i}, c_{i,i})I(c_{i,i} < v) + vI(c_{i,i} \geq v)$.

Now consider the choice of fractions $\{\alpha_i\}_{i=1,2}$. Impose symmetry: $\alpha = \alpha_1 = \alpha_2$. If $\alpha < \tilde{\alpha}$ there is no incentive to deviate: if $\alpha < \alpha^*$, deviation does not affect insurer $i$’s payoff at all (as she does not enter), and if $\alpha \geq \alpha^*$, deviations to $\alpha_i \geq \tilde{\alpha}$ simply harm insurer $i$. Thus the profiles described above are symmetric subgame perfect Nash equilibria. If $\alpha > \tilde{\alpha}$, then insurer $i$ may profitably deviate to $\alpha_i < \tilde{\alpha}$. Thus the profiles described above are the \textit{only} symmetric subgame perfect Nash equilibria.

Finally, only $\gamma = 0$ can obtain in equilibrium when $\alpha > \alpha^*$: unilateral deviations to lower $\gamma_i$ always benefit insurer $i$ if her primary market profit is positive. When $\alpha \leq \alpha^*$ the choice of $\gamma$ is irrelevant.

\textit{Proof of Proposition 4.3.} Consider deviation to $\alpha_i = \tilde{\alpha}$. Because $c_{i,i} < v$, this deviation strictly improves profits by foreclosing entry.