The promise, peril and threat of big data

Stephen J. Mildenhall

November 2015

Aon Center for Innovation and Analytics, Singapore

Data Science and the Actuary: threat or opportunity?
Donoho's six divisions of Greater Data Science

Data Exploration and Preparation
- Become one with the data: 80% of effort
- Anomalies and artifacts

Data Representation & Transformation
- Database: SQL, noSQL, distributed, live data streams
- Fourier, wavelet, multi-scale transform

Data Modeling
- Generative, inferential
- Predictive, ML

Data Visualization and Presentation
- Extends EDA
- Common task framework

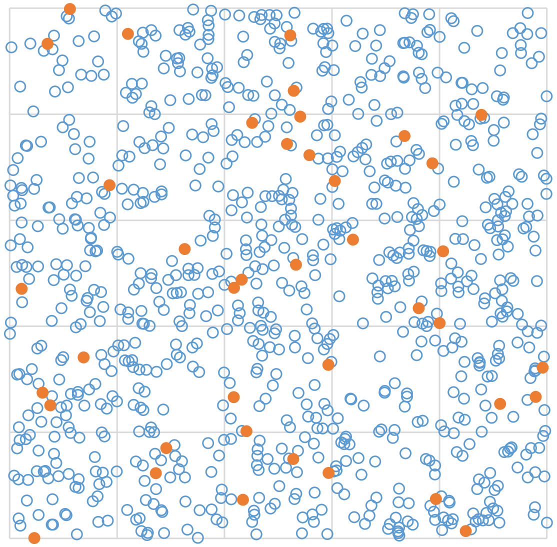
Enabled by

Computing with Data
- Important
- R, Python…be multi-lingual
- Hardware issues
- Short term

Science about Data Science

Open Science
- Closed open data
- Meta data
- Interrogating data

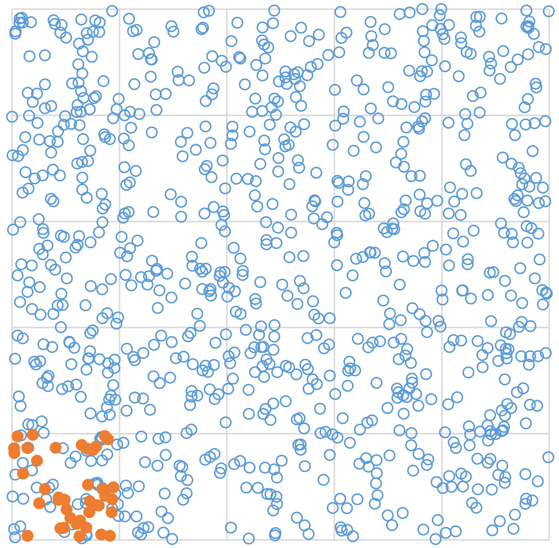
Big data and insurance: be careful what you wish for



Insurable

Old School

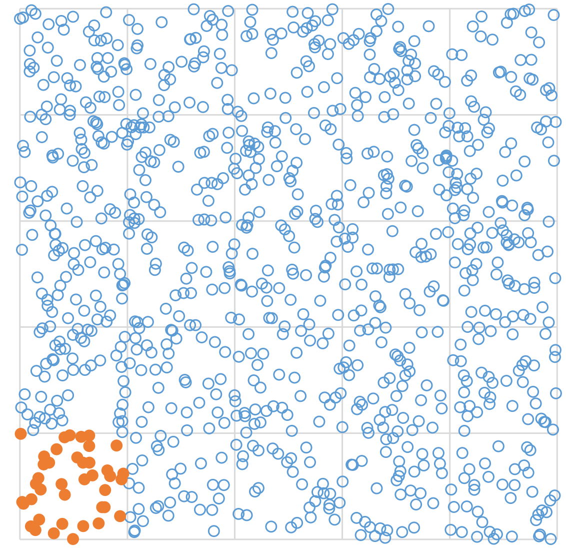
4% of 100%



Insurable, but expensive

Flood

50% of 8%



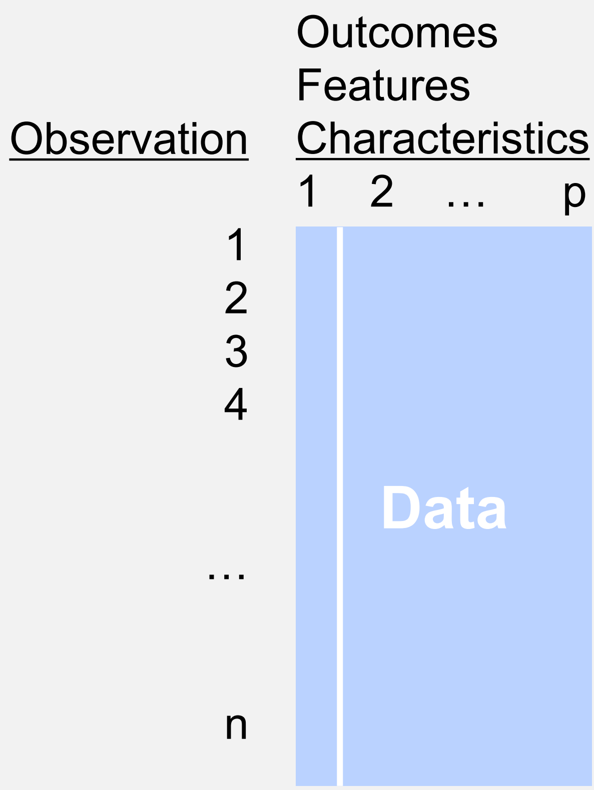
Not Insurable

Genetics

100% of 4%

What puts the Big in Big Data?

Traditional data



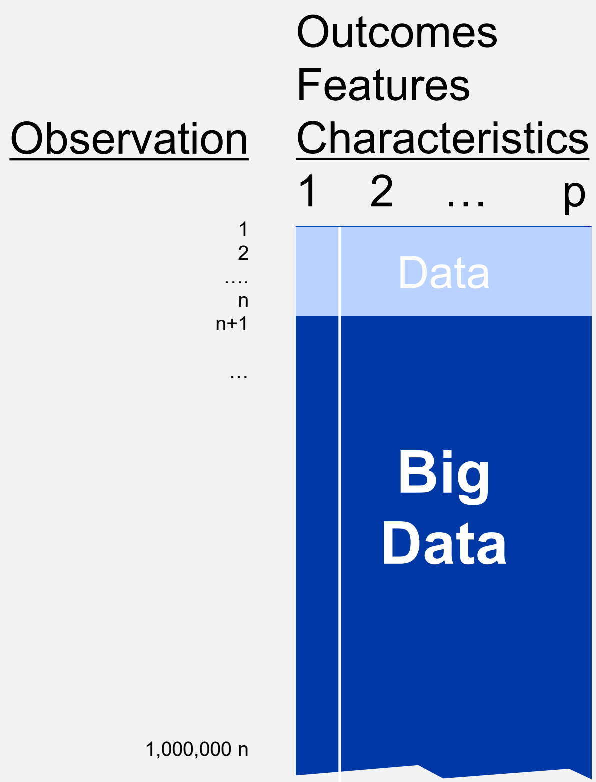
$n > p$

- Observations = insureds
- Observed quantities
  - Losses
  - Age
  - Sex
  - Marital status
  - Vehicle use
  - Accident history
  - Etc.

- Observations = sentences
- Observed quantities
  - Word frequencies

What puts the Big in Big Data?

Extension I: Lots more observations



n >> p

- Sentences, results radically improve with **billions** of test sentences
- Global satellite images
- Tick-level financial data

- Insurance examples
  - More years of **experience**
  - Some experience vs none!

- **Computing challenge**
- **Same modeling approaches**

What puts the Big in Big Data?

Extension II: Lots more parameters

Outcomes
Features
<u>Observation</u> <u>Characteristics</u>

1 2 … p … 10000 p

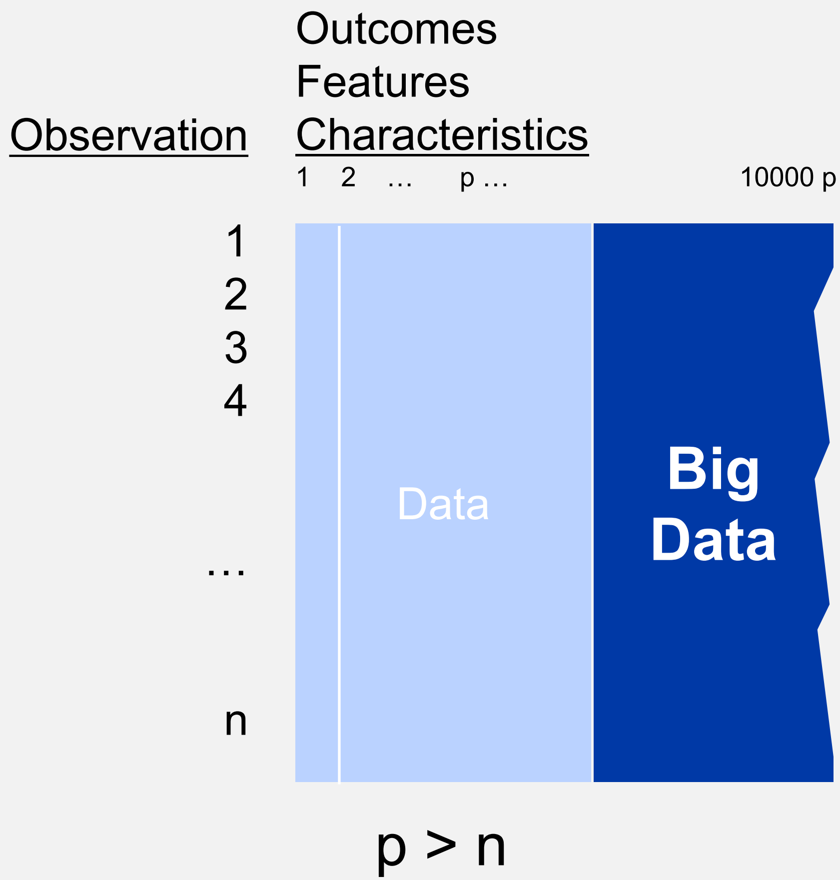
1
2
3
4
…
n

p > n

Empower Results®

What puts the Big in Big Data?

Extension II: Lots more parameters



- Detailed credit history, in use since mid-1990s
- Minute by minute driving log, auto telematics
- Home telematics
- Genome information
- Hyperspectral image
- **Computing challenge**
- **New modeling challenge**

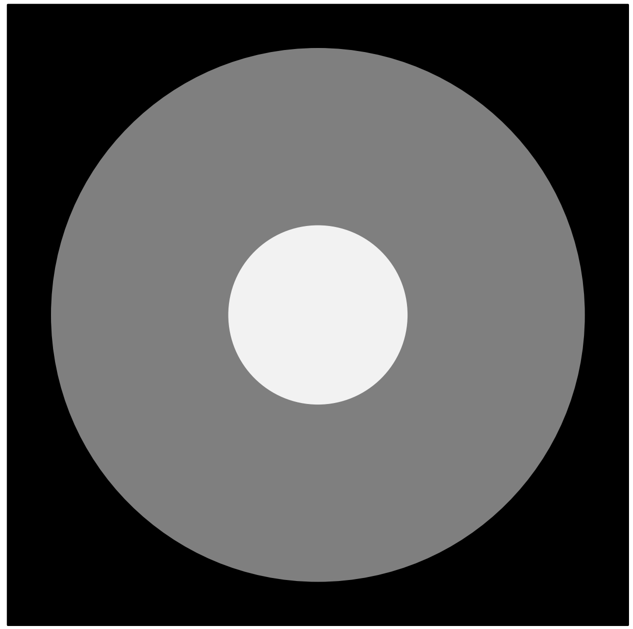
Empower Results®

How does more data impact risk and insurance?

Unknown, ignorance, no insurance

Partial knowledge, uncertainty, risk, insurance

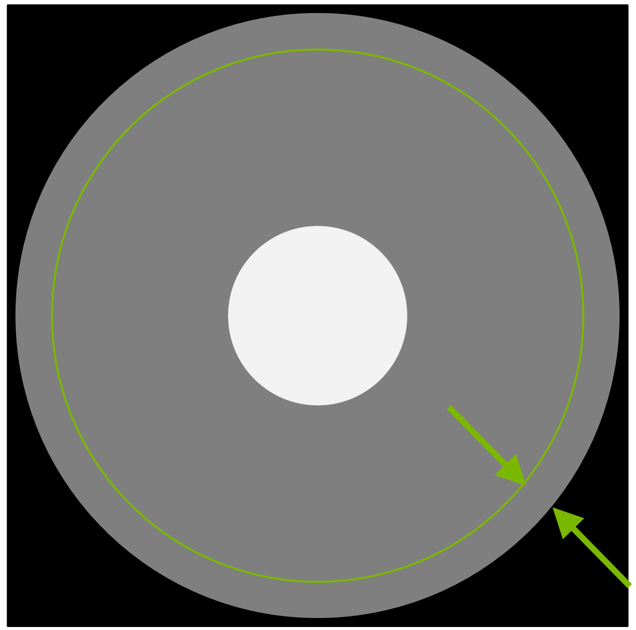
Complete knowledge, certainty, managed, retained



- **Incomplete understanding creates opportunities for insurance markets**
- Ignorance and certain knowledge generally rule out insurance

How does more data impact risk and insurance?
Extension I. More observations, bigger n

Insurance gain from decreased ignorance



More data is a GOOD THING
Emerging Risk

- **More observations can create markets**

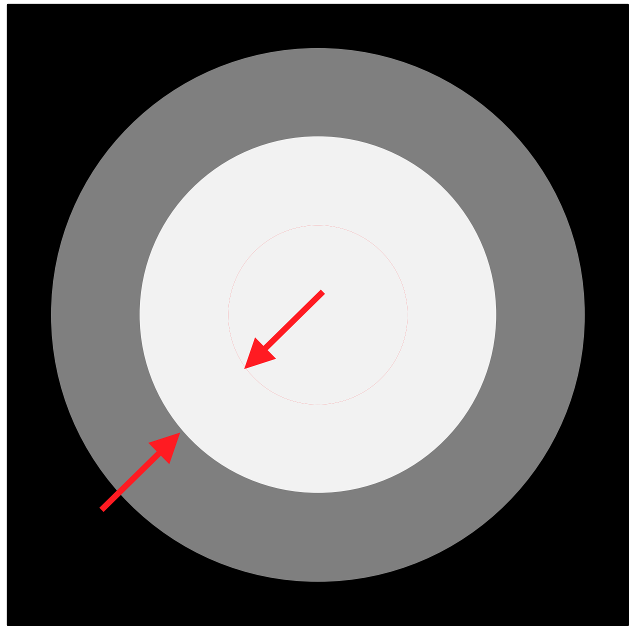
- Risk measured by risk owners
- Measurement begets management
- Risk more quantifiable for insurers

- Property catastrophe
- Cyber
- Business interruption
- Terrorism
- Giga liability
- Brand

Empower Results®

How does more data impact risk and insurance?
Extension II. More parameters, bigger p

Insurance loss from greater certainty



More data is a BAD THING
Existing Risk

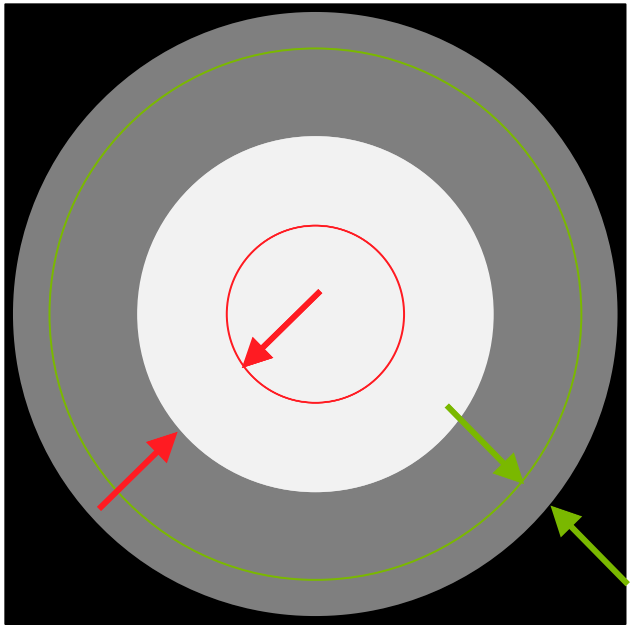
- **More parameters may destroy markets in the long run**

- More granular underwriting
- Less risk sharing
- Affordability and availability issues

- Genomics in health insurance
- Flood insurance

How does more data impact risk and insurance?

Indeterminate net growth effect



- Net growth impact on risk-transfer insurance indeterminate
- Different data models apply in different markets
- **Disruption is certain**

More data is a
???
Disruption

Empower Results®

The value of information

- Initial state of ignorance about individual risks allows market to transfer and diversify all idiosyncratic risk
  – First-best outcome is for agents to fully insure their risk at the actuarially fair premium
  – Best because of risk aversion

- For the same reason, risk aversion, information always has a nonnegative value for the decision maker…
  – Adding information lowers the variance of the outcome distribution, X → (X | information) has a “less risky” distribution

- **…assuming the information does not affect the other parameters of the environment for the decision maker**
  – The information is private and not public

Empower Results®

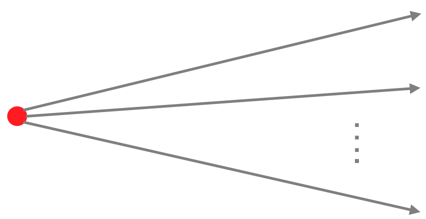
Insurance market with no individual risk information

Ignorance → **Losses Revealed**

- Underwriting and risk management decisions made in state of minimal specific risk information

- Actual losses revealed to insurer and insured ***after*** policy contracting

Range of outcomes

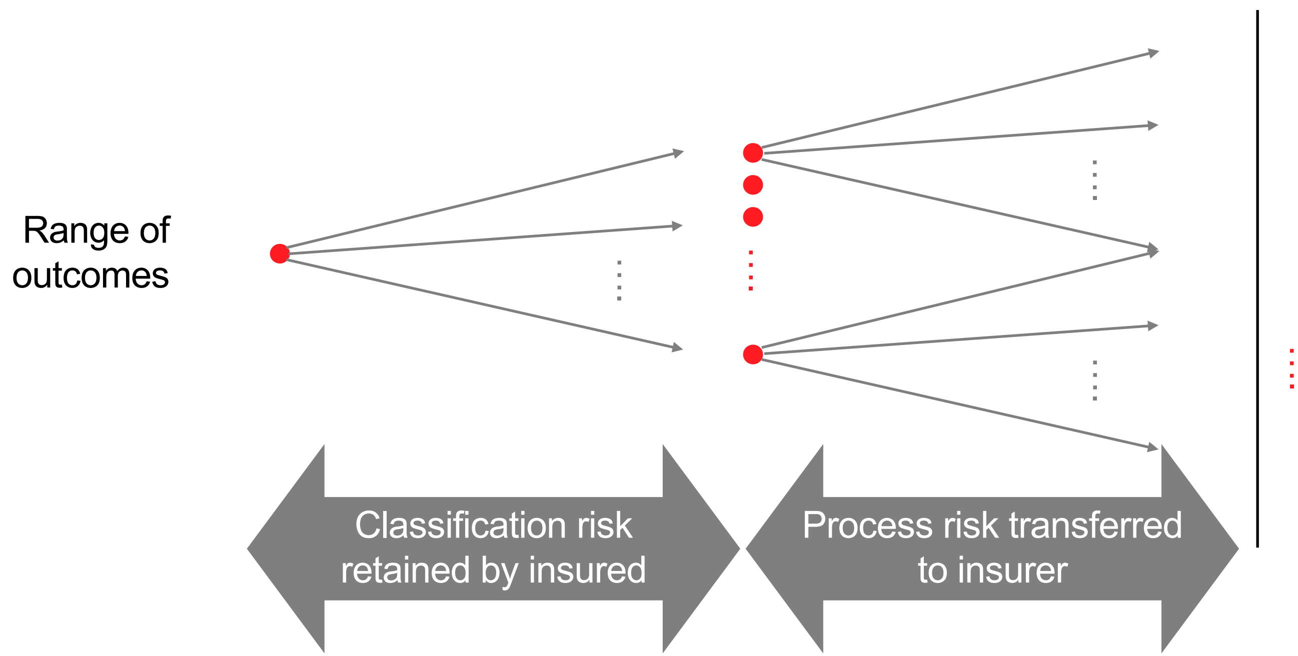


Classification and process risk transferred

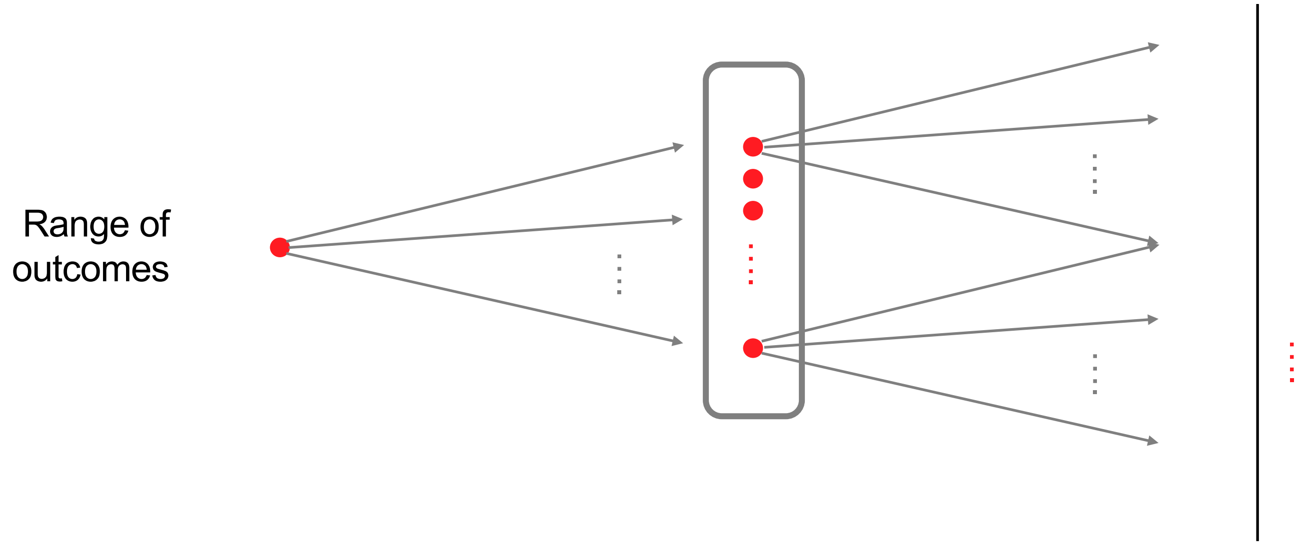
Empower Results®

Hirshleifer Effect: classification risk becomes uninsurable… making everyone worse off

Ignorance	Partial Knowledge Detailed u/w class	Losses Revealed
▪ Detailed information about insureds not available before policy contracting	▪ **Insured bears classification risk, which has now become uninsurable**	▪ Insured loss distribution revealed conditional on detailed under-writing information



Link to Buhlmann greatest accuracy credibility theory



Variance of hypothetical means

- Spread of targets
- VHM

Expected value of process variance

- Accuracy of shot
- EVPV

$$Z = \frac{n}{n + K}$$

$$K = \frac{EVPV}{VHM}$$

- Classes of business with a relatively higher VHM that is captured by the classification scheme, and lower EVPV (low K, high credibility) have more to lose than classes with low credibility
- Higher frequency, lower severity classes most at risk
- Flood is a good example of a high risk class

Empower Results®

Possible remedies for the Hirshleifer effect

Organize insurance before information becomes available

- Long-term contracts can provide re-classification risk, e.g. whole life, health, LTC
- Hard to guarantee no one has information when contract executed, adverse selection
- Cancellation problem: those with good information cancel

Ban information technology

- Hard to organize, can't uninvent technologies
- Counterproductive, e.g. medical tests needed to ensure delivery of best treatment
- Prohibiting use in u/w leads to adverse selection problem with asymmetric information

Socialize risk through compulsory insurance

- Social security
- Private market solutions have problem of "buying a loss" leading to need for residual market mechanisms to be insurers of last resort

Unraveling due to adverse selection

North American Actuarial Journal, 19(1), 60–72, 2015
Copyright © Society of Actuaries
ISSN: 1092-0277 print / 2325-0453 online
DOI: 10.1080/10920277.2014.982871

Routledge
Taylor & Francis Group

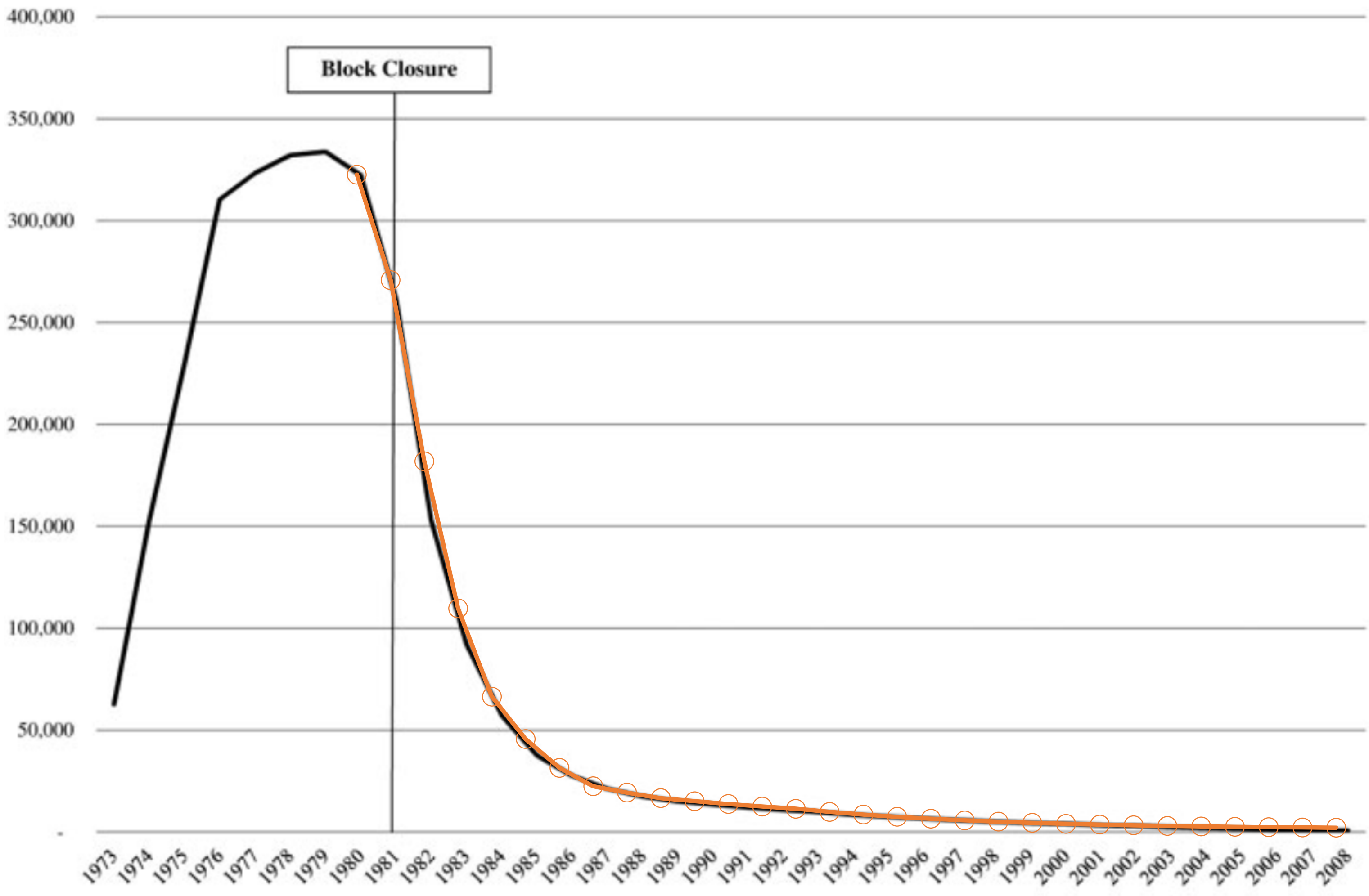
Anatomy of a Slow-Motion Health Insurance Death Spiral

H. E. Frech III[1] and Michael P. Smith[2]
[1]*Department of Economics, University of California, Santa Barbara, California*
[2]*Compass Lexecon, Los Angeles, California*

Adverse selection death spirals in health insurance are dramatic and, so far, exotic economic events. The possibility of death spirals has garnered recent policy and popular attention because the pricing regulations in the Affordable Care Act (ACA) of 2010 make health plans more vulnerable to them (though some other aspects of the ACA limit them). Most death spirals tracked in the literature have involved selection against a group health plan that was dropped quickly by the employer. In this article, we empirically document a death spiral in individual health insurance that was apparently triggered by a block closure in 1981 and developed slowly because the insurer partially subsidized the block. We show that premiums rose dramatically from around the time of the block closure to at least 2009 (the last year of available data). By 2009, some, but very few, policyholders remained in the block, and premiums were roughly seven times that of a yardstick we developed. The history of this slow-moving event is directly relevant to current policy discussions because of both adverse selection in general and the particular problems induced by closing a block.

Empower Results®

Insureds run-off, 1980 to 2009
Organic neural network fit



Details of lognormal model

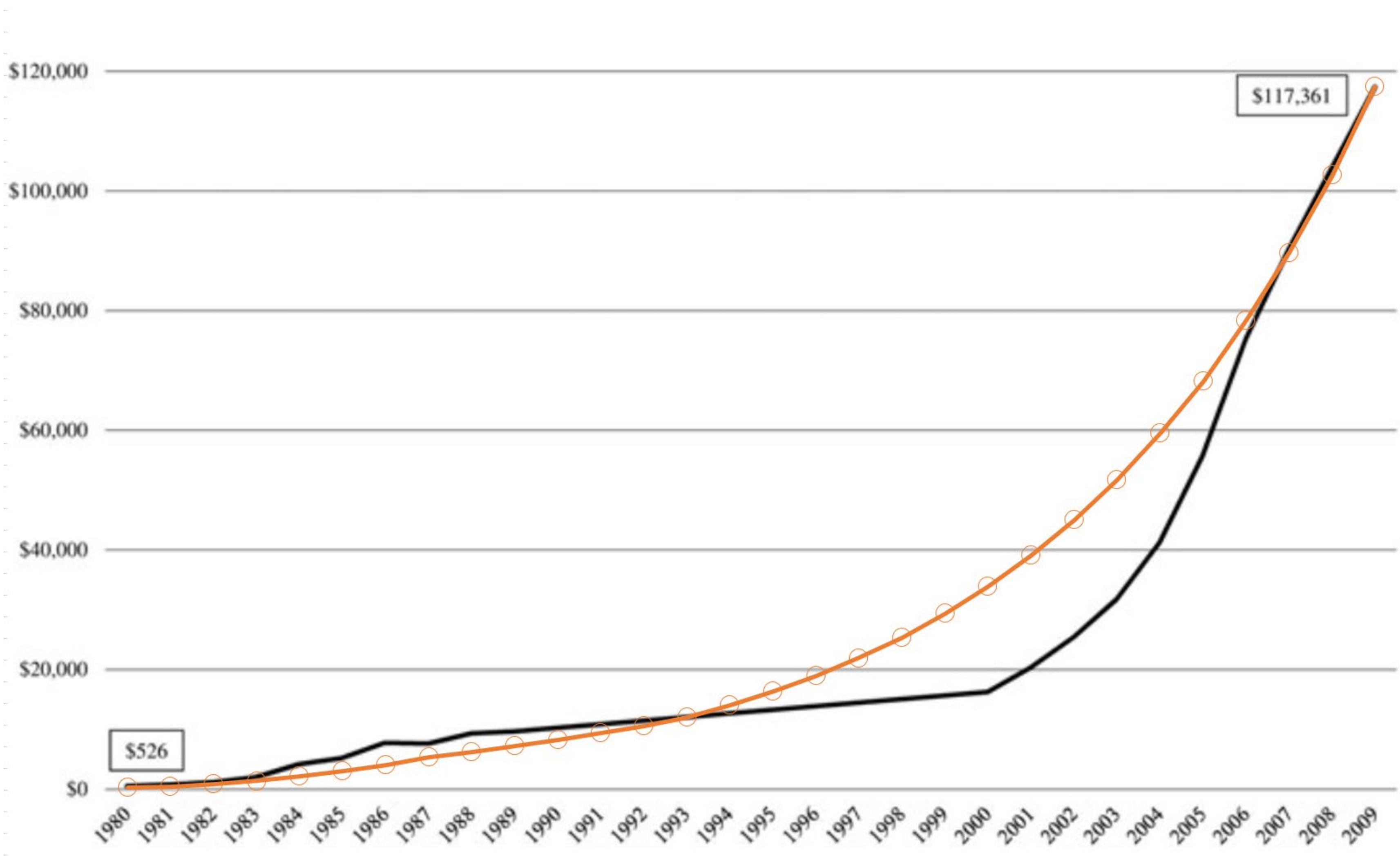
Lognormal Model of Market Unraveling, implied CV = 2.06

Year	Mu with 7.5% Medical Inflation	Adverse Selection Threshold	F(x)	Expense(x)	Conditional Value = Premium	Rate Change	Insureds	Insured Fall off Factor	Implied 1-F
1980	5.438	0	0.0000	526	526		325,000	1.00	1.000
1981	5.510	69	0.1613	501	666	26.7%	272,581	0.84	0.839
1982	5.582	217	0.4381	440	1,001	50.2%	182,629	0.67	0.562
1983	5.655	490	0.6628	362	1,564	56.2%	109,577	0.60	0.337
1984	5.727	897	0.7977	292	2,343	49.8%	65,746	0.60	0.202
1985	5.799	1,344	0.8624	251	3,169	35.3%	44,708	0.68	0.138
1986	5.872	1,937	0.9065	214	4,226	33.3%	30,401	0.68	0.094
1987	5.944	2,663	0.9345	185	5,483	29.8%	21,281	0.70	0.065
1988	6.016	3,181	0.9443	179	6,403	16.8%	18,089	0.85	0.056
1989	6.089	3,786	0.9527	174	7,463	16.6%	15,375	0.85	0.047
1990	6.161	4,340	0.9574	175	8,447	13.2%	13,838	0.90	0.043
1991	6.233	4,969	0.9617	176	9,555	13.1%	12,454	0.90	0.038
1992	6.306	5,683	0.9655	177	10,801	13.0%	11,209	0.90	0.034
1993	6.378	6,491	0.9690	177	12,202	13.0%	10,088	0.90	0.031
1994	6.450	7,672	0.9738	171	14,182	16.2%	8,524	0.84	0.026
1995	6.523	9,043	0.9778	164	16,460	16.1%	7,202	0.84	0.022
1996	6.595	10,635	0.9813	158	19,077	15.9%	6,086	0.84	0.019
1997	6.667	12,478	0.9842	152	22,081	15.7%	5,142	0.84	0.016
1998	6.740	14,611	0.9866	146	25,525	15.6%	4,345	0.84	0.013
1999	6.812	17,076	0.9887	140	29,472	15.5%	3,671	0.84	0.011
2000	6.884	19,920	0.9905	134	33,989	15.3%	3,102	0.84	0.010
2001	6.956	23,197	0.9919	129	39,156	15.2%	2,621	0.84	0.008
2002	7.029	26,970	0.9932	123	45,061	15.1%	2,215	0.84	0.007
2003	7.101	31,309	0.9942	118	51,805	15.0%	1,871	0.84	0.006
2004	7.173	36,294	0.9951	113	59,501	14.9%	1,581	0.84	0.005
2005	7.246	42,015	0.9959	108	68,276	14.7%	1,336	0.84	0.004
2006	7.318	48,575	0.9965	103	78,276	14.6%	1,129	0.84	0.003
2007	7.390	56,089	0.9971	99	89,663	14.5%	954	0.84	0.003
2008	7.463	64,689	0.9975	94	102,622	14.5%	806	0.84	0.002
2009	7.535	74,524	0.9979	90	117,361	14.4%	681	0.84	0.002

Empower Results®

Lognormal severity, 7.5% trend, fit to starting & ending values

Rate change suppressed during experience period; includes medical trend and age effect, 32 to 61



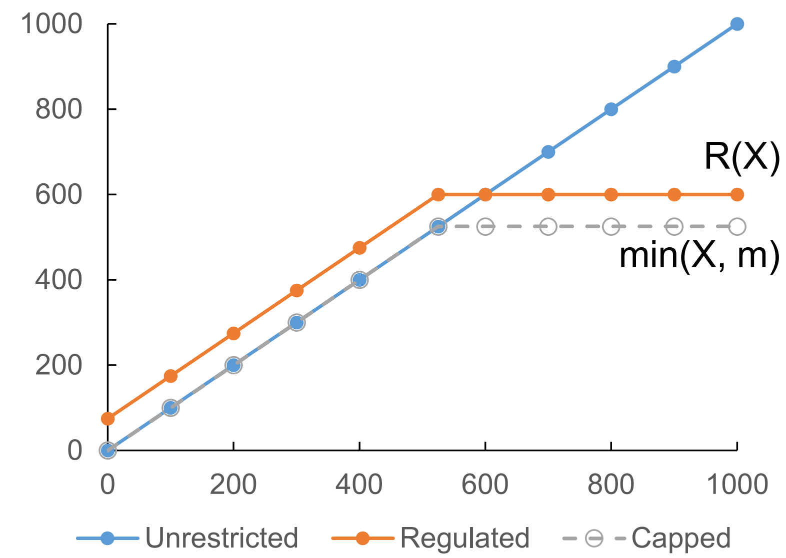
Implied percentiles of spend distribution
1980 and 2009 Dollars

Percentiles of Severity Distribution

Percentile	1980	2009	Plus Attrition
50.0%	230	1,873	98,009
90.0%	1,196	9,738	175,175
95.0%	1,908	15,540	220,669
99.0%	4,585	37,343	364,207
99.853%	12,251	99,773	697,422

Regulator caps premiums to ensure affordability

- Model premium as R(X) = M – m + min(X, m)
  - M = maximum premium charged to any insured
  - m = maximum "variable" premium, related to individual risk
  - M – m represents the market access fee or residual market load needed to ensure solvency
  - M is a "policy" variable = set by policy maker
  - m is set by the constraint E(X) = E(R(X))
    - m solves the equation M – m = Expense(m)
    - E(min(X,m)) + Expense(m) = E(X), so E(R(X)) = M – m + E(X) – Expense(m) = E(X)



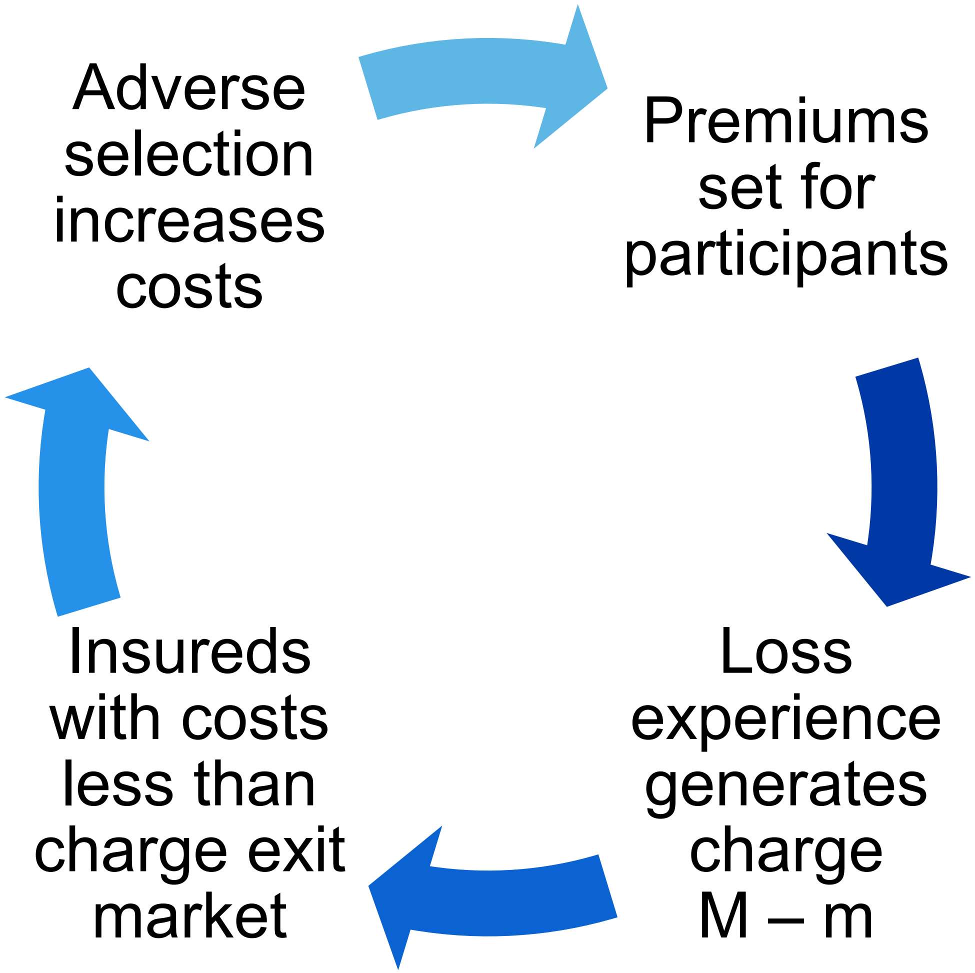
M = 600
m = 525
Charge = M – m = 75

- For simplicity assume insureds "opt out" if R(X) is more than double their (known) loss cost X, i.e. opt out if X < (M – m) the "market access fee"

- **What is the critical level of CV where market starts to unravel?**

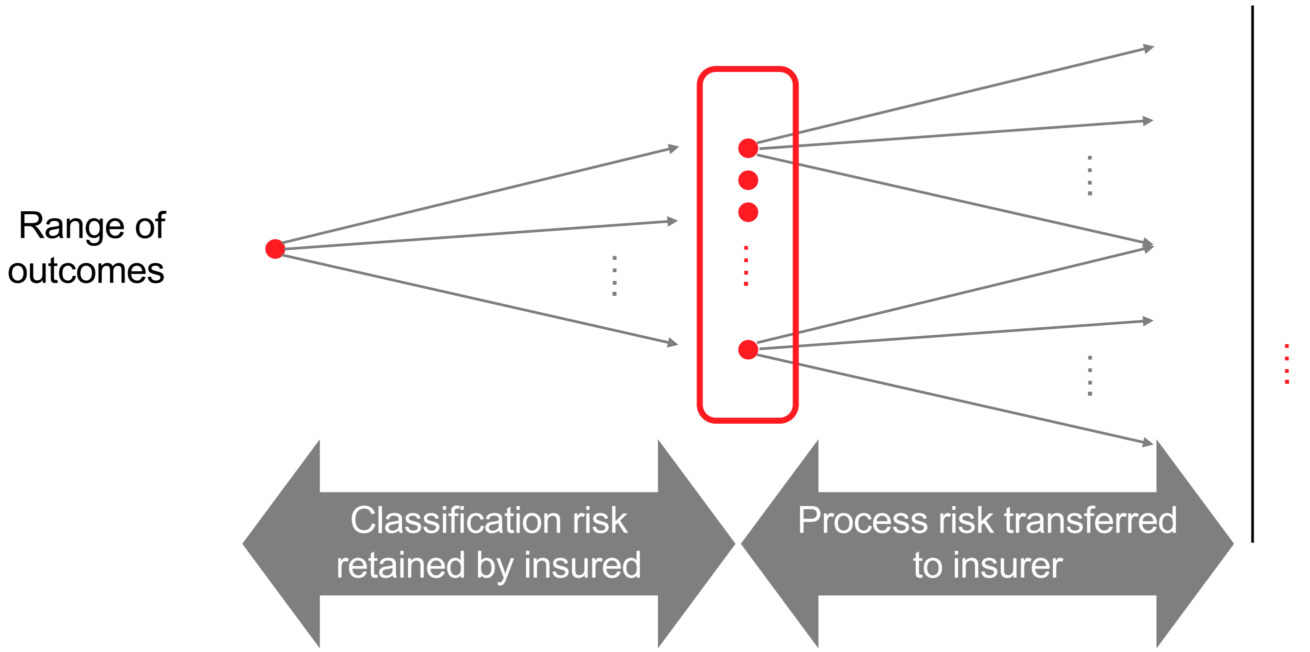
Empower Results®

Market dynamics



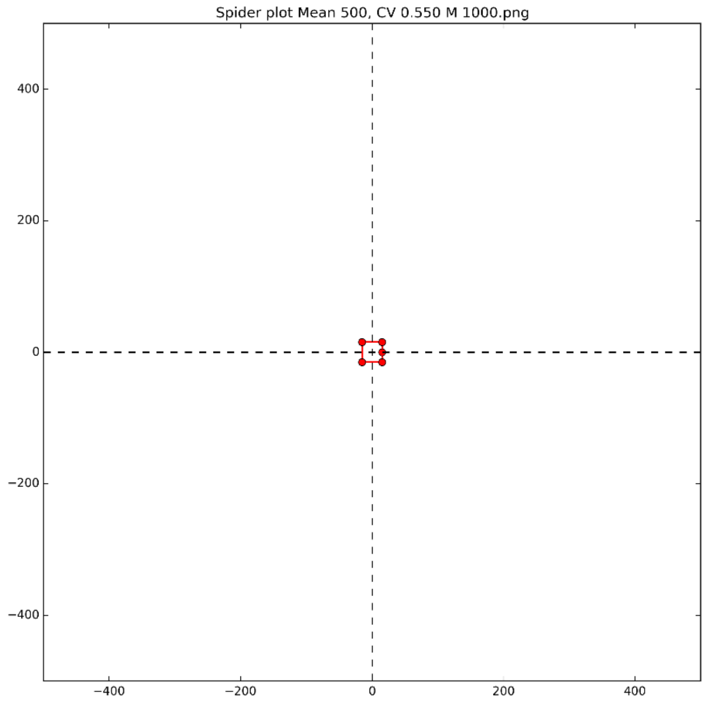
Key variable = coefficient of variation (CV) of rating plan

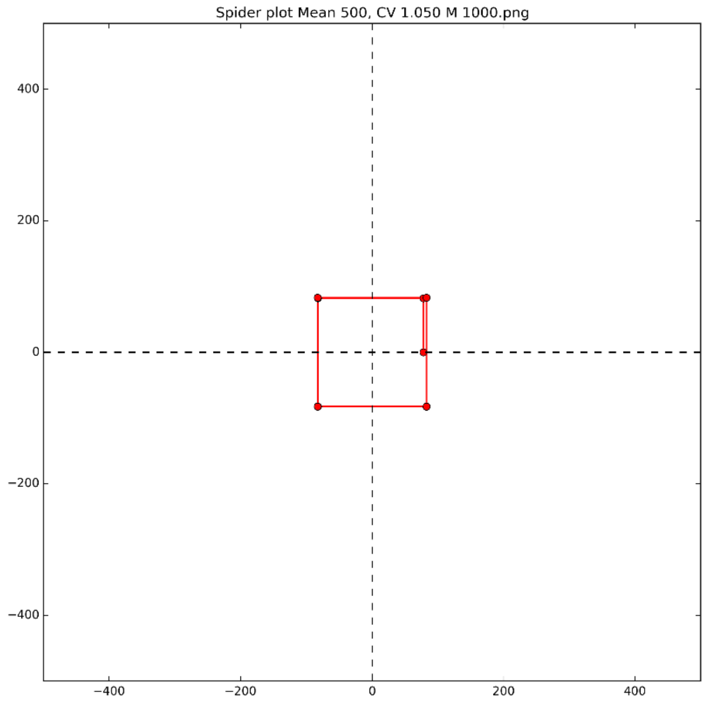
- Key variable = variance of hypothetical means
- Essentially the variance of the premiums in the classification plan



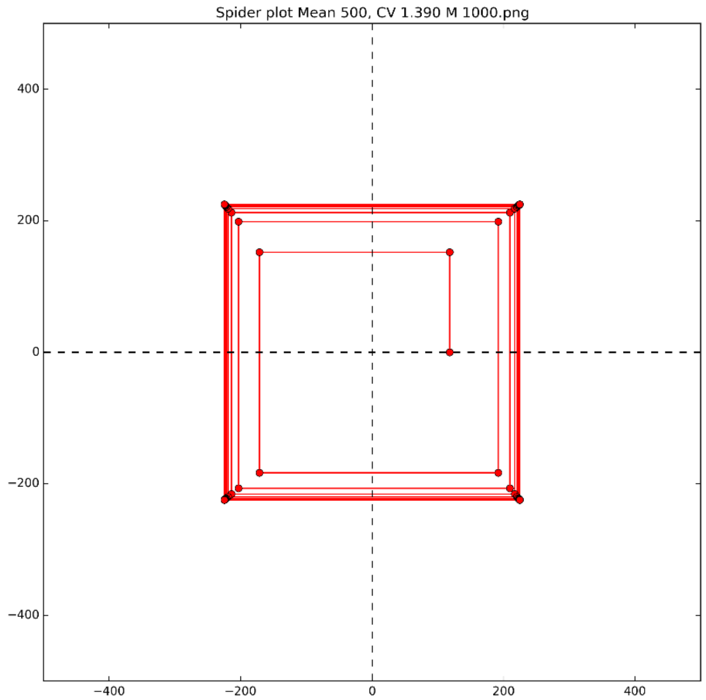
Aon Empower Results®

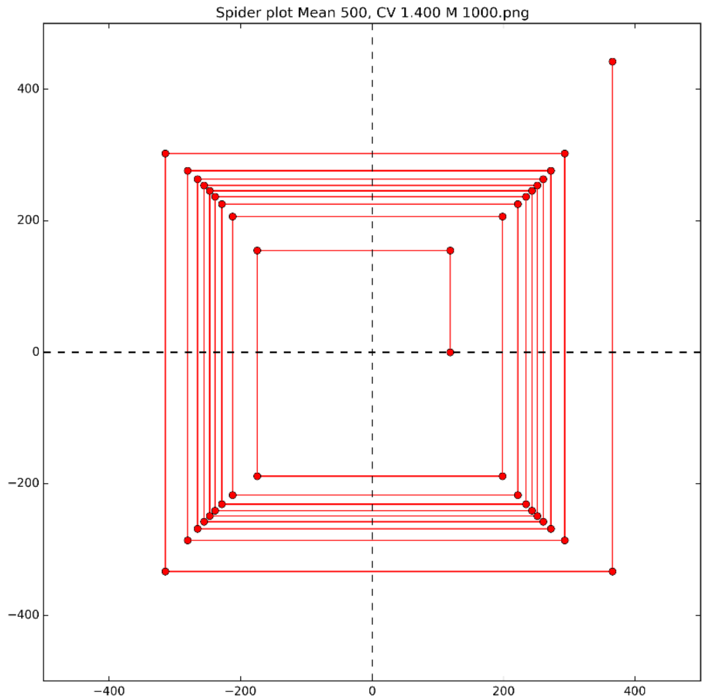
For low to moderate CVs market is stable and charge small



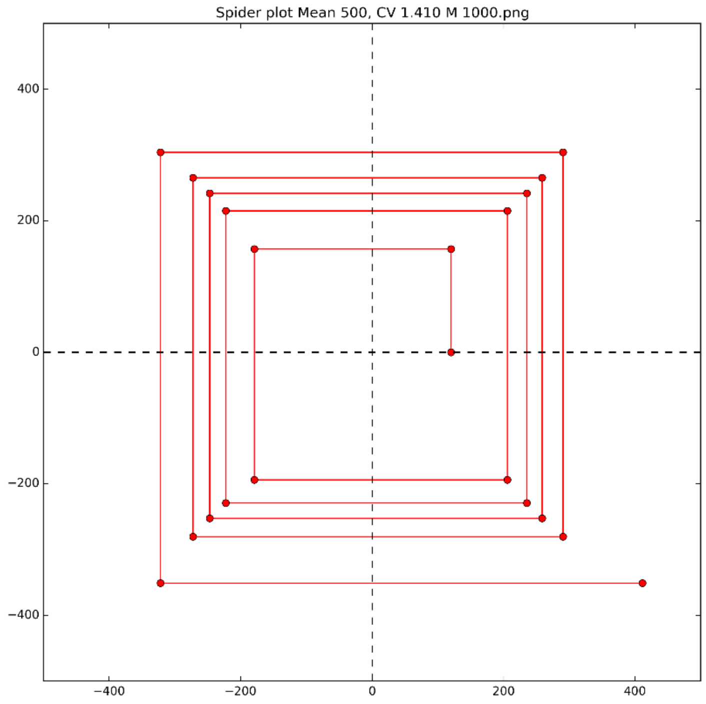


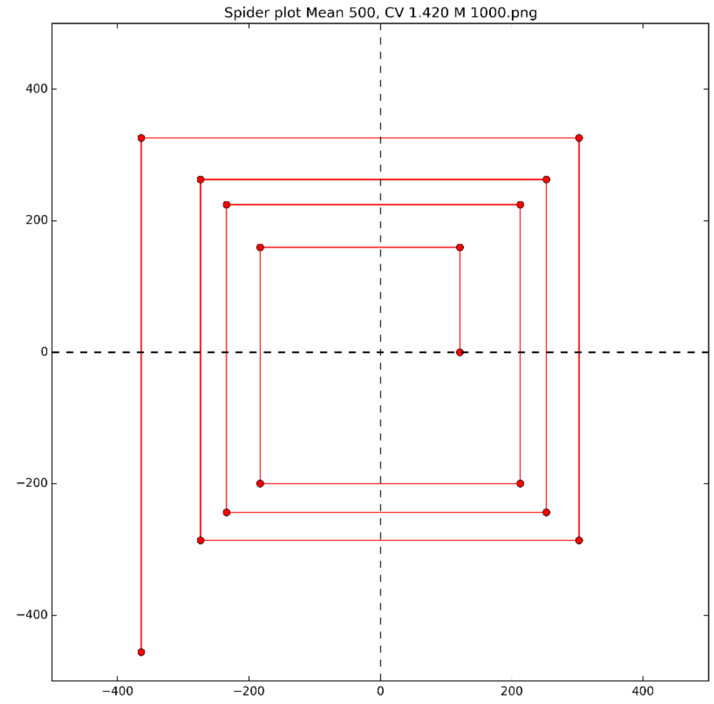
But instability develops…





…above a critical level





Empower Results®

What about real world examples?

- Auto data
  - CV of hypothetical means, i.e. classification means, is around 35%
  - **Stable behavior** of simple model for volatility in the same range

- Flood data
  - CV of hypothetical means, i.e. classification means, is around 1.50 based on Aon Benfield Impact Forecasting riverine flood model and a sample of notional risks
  - Highly skewed distribution; many risks have very low expected losses
  - **Unstable behavior** of simple model for volatility in the same range

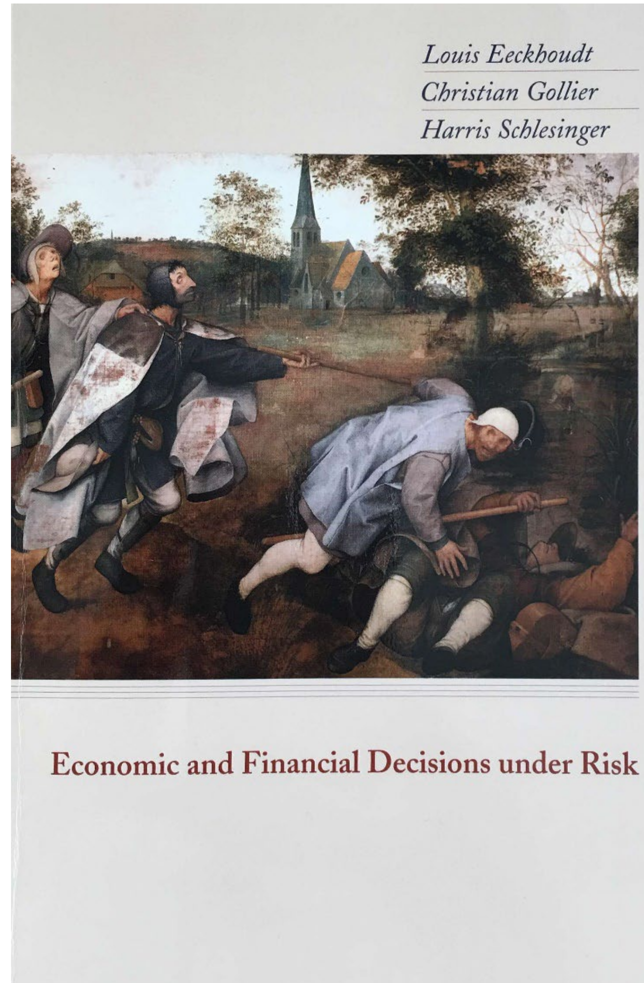
References

The Private and Social Value of Information and the Reward to Inventive Activity

By Jack Hirshleifer*

A number of recent papers[1] have dealt with the economics of information in a context in which each individual is fully certain about his own endowment and productive opportunities. In those papers, the individual is imperfectly informed only about his market opportunities, i.e., about the supply-demand offers of *other* individuals. In consequence, costly patterns of search for trading partners replace the traditional assumption of costless exchange.

Technological uncertainty brings immediately to mind the economics of research and invention. The traditional position has been that the excess of the social over the private value of new technological knowledge leads to underinvestment in inventive activity. The main reason is that information, viewed as a product, is only imperfectly appropriable by its discoverer.[3] But this paper will show that there is a hitherto unrecognized force operating in the opposite direction. What has been scarcely



Contact information

Stephen Mildenhall
Chairman of Aon Center for Analytics and Innovation
Chief Executive Officer, Aon Benfield Analytics
stephen.mildenhall@aon.com
Singapore +65 6872 7668
US +1 312 381 5880

Stephen Mildenhall is Chairman of the Aon Center for Innovation and Analytics in Singapore, which leverages Aon's data assets to provide analytically driven solutions for clients across all of Aon's businesses. He is also CEO of Aon Benfield Analytics, a global team of over 500 professionals dedicated to helping clients measure, monitor, manage and profit from risks within their portfolios.

Steve joined Aon in 2003. Prior to Aon, he held various actuarial positions at Kemper Insurance, CNA Re Facultative and CNA Personal Lines, all in Chicago. He started in the insurance industry in 1992.

Steve is a Fellow of the Casualty Actuarial Society, an Associate of the Society of Actuaries and a Member of the American Academy of Actuaries. He is a Chartered Enterprise Risk Analyst. He received his Masters and PhD degrees in Mathematics from the University of Chicago, and a BSc in Mathematics from the University of Warwick in England.

Empower Results®