

No Free Lunch for Stress Testers

Toward a Normative Theory of Scenario Based Risk Assessment

Roger M. Stein

Lisa R. Goldberg

Modern Massive Data Sets

MIT Laboratory for Financial Engineering
Center for Risk Management Research



June 20, 2014

DRAFT: PLEASE DO NOT QUOTE OR REPRODUCE
WITHOUT CONSENT OF THE AUTHORS

The Financial Crisis of 2008–09

The US stock market lost 40% of its value



The Financial Crisis of 2008–09

The US stock market lost 40% of its value
[Atkinson et al., 2013] estimate the damage:



The Financial Crisis of 2008–09

The US stock market lost 40% of its value

[Atkinson et al., 2013] estimate the damage:

- ▶ Between \$6 and \$14 trillion (between \$50 and \$120 thousand per household) in GDP foregone



The Financial Crisis of 2008–09

The US stock market lost 40% of its value

[Atkinson et al., 2013] estimate the damage:

- ▶ Between \$6 and \$14 trillion (between \$50 and \$120 thousand per household) in GDP foregone
- ▶ Household net worth fell by \$16 trillion (24%) between Q3 2007 and Q1 2009



The Financial Crisis of 2008–09

The US stock market lost 40% of its value

[Atkinson et al., 2013] estimate the damage:

- ▶ Between \$6 and \$14 trillion (between \$50 and \$120 thousand per household) in GDP foregone
- ▶ Household net worth fell by \$16 trillion (24%) between Q3 2007 and Q1 2009
- ▶ Peak unemployment rate of 10%



The Financial Crisis of 2008–09

The US stock market lost 40% of its value

[Atkinson et al., 2013] estimate the damage:

- ▶ Between \$6 and \$14 trillion (between \$50 and \$120 thousand per household) in GDP foregone
- ▶ Household net worth fell by \$16 trillion (24%) between Q3 2007 and Q1 2009
- ▶ Peak unemployment rate of 10%
- ▶ Median duration unemployment at 25 weeks (against a benchmark of 5-10 weeks)



The Dodd-Frank Wall Street Reform and Consumer Protection Act was signed into law in 2010; its mission is:

To promote the financial stability of the United States by improving accountability and transparency in the financial system, to end “too big to fail,” to protect the American taxpayer by ending bailouts, to protect consumers from abusive financial services practices, and for other purposes



The Dodd-Frank Wall Street Reform and Consumer Protection Act was signed into law in 2010; its mission is:

To promote the financial stability of the United States by improving accountability and transparency in the financial system, to end “too big to fail,” to protect the American taxpayer by ending bailouts, to protect consumers from abusive financial services practices, and for other purposes

Stress tests are an integral part of Dodd-Frank regulation



Financial Stress Tests

Used to assess capital adequacy: will an important financial institution get into trouble in a market downturn?



Financial Stress Tests

Used to assess capital adequacy: will an important financial institution get into trouble in a market downturn?

Conceptually simpler than statistical risk measures, such as volatility, value at risk and expected shortfall



Are Stress Tests Effective?



Stress Testing as a Classification Problem

In a rudimentary framework, a stress tester tasked with determining capital adequacy assigns a grade of OK (1) or Trouble (0) to each portfolio.



Stress Testing as a Classification Problem

In a rudimentary framework, a stress tester tasked with determining capital adequacy assigns a grade of OK (1) or Trouble (0) to each portfolio.

The output of a stress test is a binary string whose length, N , is the number of portfolios



Stress Testing as a Classification Problem

In a rudimentary framework, a stress tester tasked with determining capital adequacy assigns a grade of OK (1) or Trouble (0) to each portfolio.

The output of a stress test is a binary string whose length, N , is the number of portfolios

There are 2^N such strings



A Simple World

There are exactly two portfolios: one with treasury bonds, one with high yield bonds

Rule	Action
s_1	Both-OK
s_2	Both-Trouble
s_3	High Yield-OK-Treasury-Trouble
s_4	Treasury-OK-High Yield-Trouble



A Simple World

There are exactly two portfolios: one with treasury bonds, one with high yield bonds

Rule	Action
s_1	Both-OK
s_2	Both-Trouble
s_3	High Yield-OK-Treasury-Trouble
s_4	Treasury-OK-High Yield-Trouble

This means that there are (for all intents and purposes) exactly $4 = 2^2$ stress tests s



A Simple World

There are exactly two portfolios: one with treasury bonds, one with high yield bonds

Rule	Action
s_1	Both-OK
s_2	Both-Trouble
s_3	High Yield-OK-Treasury-Trouble
s_4	Treasury-OK-High Yield-Trouble

This means that there are (for all intents and purposes) exactly $4 = 2^2$ stress tests s

Which stress test is best?



The Future in a Simple World

Portfolio	f_1	f_2	f_3	f_4
High Yield	OK	Trouble	OK	Trouble
Treasury	OK	Trouble	Trouble	OK

There are also exactly four possible future outcomes f



Stress Testers Make Errors

The error made by a stress test s on portfolio x under future outcome f :

$$\varepsilon(s | x, f) = |s(x) - f(x)|. \quad (1)$$



Stress Testers Make Errors

The error made by a stress test s on portfolio x under future outcome f :

$$\varepsilon(s | x, f) = |s(x) - f(x)|. \quad (1)$$

The aggregate error made by a stress test s under future outcome f :

$$\mathcal{E}(s | f) = \sum_x \varepsilon(s | x, f) P(x | f) \quad (2)$$

The OTS error of a stress test s is:

$$\mathcal{E}(s) = \sum_f \mathcal{E}(s | f) P(f) \quad (3)$$



No Free Lunch for Stress Testers in a Simple World

Aggregate errors made by a stress tester if all portfolios x are equally important

	f_1	f_2	f_3	f_4
s_1	0	2	1	1
s_2	2	0	1	1
s_3	1	1	0	2
s_4	1	1	2	0



No Free Lunch for Stress Testers in a Simple World

Aggregate errors made by a stress tester if all portfolios x are equally important

	f_1	f_2	f_3	f_4
s_1	0	2	1	1
s_2	2	0	1	1
s_3	1	1	0	2
s_4	1	1	2	0

The row averages are all the same



No Free Lunch for Stress Testers in a Simple World

Aggregate errors made by a stress tester if all portfolios x are equally important

	f_1	f_2	f_3	f_4
s_1	0	2	1	1
s_2	2	0	1	1
s_3	1	1	0	2
s_4	1	1	2	0

The row averages are all the same

They represent OTS errors $\mathcal{E}(s)$ assuming the states f are equally probable



No Free Lunch for Computer Scientists

The equality of the row averages is no accident



No Free Lunch for Computer Scientists

The equality of the row averages is no accident

It is a consequence of the No Free Lunch Theorems of computer science



No Free Lunch for Computer Scientists

The equality of the row averages is no accident

It is a consequence of the No Free Lunch Theorems of computer science

The row averages (OTS Errors) **must** be the same of the states f are equally probable; more information is in [Wolpert and Macready, 1997] and [Lattimore and Hutter, 2013]



Informed Probabilities in a Simple World

	f_1	f_2	f_3	f_4
High Yield	OK	Trouble	OK	Trouble
Treasury	OK	Trouble	Trouble	OK
Equal Probabilities	.25	.25	.25	.25
Informed Probabilities	.19	.19	.02	.60

A stress test designer implicitly incorporates his or her views in the design of a stress test by emphasizing the most probable bad outcomes and the most important portfolios



OTS Errors in a Simple World

	with Equal Probabilities	with Informed Probabilities
s_1	1.00	1.00
s_2	1.00	1.00
s_3	1.00	1.58
s_4	1.00	0.42

Rule s_4 OKs the treasury portfolio but flags the high yield portfolio as potentially troublesome; rule s_3 reverses the designations



Still No Free Lunch for Stress Testers in a Simple World...

...though this time it is the designer of the rules that must pay:

- ▶ Either the stress tester must admit that no rule will perform better than another in expectation
- ▶ Or he or she must explicitly state assumptions about the distribution of outcomes (f) that allow assert a rule has predictive OTS error



Toward a Normative Theory of Scenario Based Risk Assessment

A stress test designer implicitly incorporates his or her views in the design of a stress test



Toward a Normative Theory of Scenario Based Risk Assessment

A stress test designer implicitly incorporates his or her views in the design of a stress test

A stress test evaluator explicitly incorporates his or her views by weighting the most important portfolios and most probable bad outcomes in out-of-sample tests



Toward a Normative Theory of Scenario Based Risk Assessment

A stress test designer implicitly incorporates his or her views in the design of a stress test

A stress test evaluator explicitly incorporates his or her views by weighting the most important portfolios and most probable bad outcomes in out-of-sample tests

It is a challenge to get this right since there are 2^N outcomes



Toward a Normative Theory of Scenario Based Risk Assessment

A stress test designer implicitly incorporates his or her views in the design of a stress test

A stress test evaluator explicitly incorporates his or her views by weighting the most important portfolios and most probable bad outcomes in out-of-sample tests

It is a challenge to get this right since there are 2^N outcomes

Factor models substantially reduce the size of the problem, but it is still enormous



Toward a Normative Theory of Scenario Based Risk Assessment

A stress test designer implicitly incorporates his or her views in the design of a stress test

A stress test evaluator explicitly incorporates his or her views by weighting the most important portfolios and most probable bad outcomes in out-of-sample tests

It is a challenge to get this right since there are 2^N outcomes

Factor models substantially reduce the size of the problem, but it is still enormous

More information is in [Stein, 2012] and [Stein and Goldberg, 2014]



- ▶ Stress-testing can be framed as a classification problem



Summary

- ▶ Stress-testing can be framed as a classification problem
- ▶ The NFL Theorems provide a useful descriptive (and to a lesser degree normative) framework for stress test design and interpretation



- ▶ Stress-testing can be framed as a classification problem
- ▶ The NFL Theorems provide a useful descriptive (and to a lesser degree normative) framework for stress test design and interpretation



Summary (continued)

- ▶ Without information about the future of the economy, all stress tests have the same average performance



Summary (continued)

- ▶ Without information about the future of the economy, all stress tests have the same average performance
- ▶ There is no free lunch: the value of passing a stress test is directly proportional to the predictive power of the stress tester's priors
 - ▶ If priors are crisp, passing a stress test has a direct interpretation as a measure of solvency
 - ▶ If priors are soft, passing a stress test has at best directional information: it is neither a necessary or sufficient condition

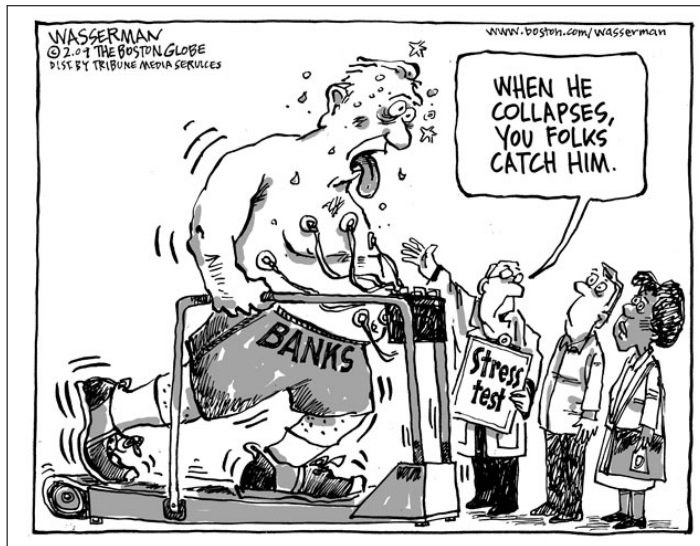


Summary (continued)

Because the space of future possibilities is so large, the performance under a stressful scenario does not generalize across portfolios, time, or economic states of the world



And What if We Don't Get This Right?



References I

[Atkinson et al., 2013] Atkinson, T., Luttrell, D., and Rosenblum, H. (2013).

How bad was it? the costs and consequences of the 2007–09 financial crisis.

DallasFed Staff Paper No. 20.

[Lattimore and Hutter, 2013] Lattimore, T. and Hutter, M. (2013).

No free lunch versus occams razor in supervised learning.

In Dowe, D. L., editor, *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence*, volume 7070, pages 223–235. Springer.

[Stein, 2012] Stein, R. M. (2012).

The role of stress testing in credit risk management.

Journal of Investment Management, 10(4):64–90.



[Stein and Goldberg, 2014] Stein, R. M. and Goldberg, L. R. (2014).

No free lunch for stress testers: Toward a normative theory of scenario based risk assessment.

manuscript in progress.

[Wolpert and Macready, 1997] Wolpert, D. H. and Macready, W. G. (1997).

No free lunch theorem for optimization.

IEEE Transactions on Evolutionary Computation, 1(1):67–82.

