Modeling the Economics of Long-Term Storage

David S. H. Rosenthal

LOCKSS Program
Stanford University Libraries
http://www.lockss.org/

LOTS OF COPIES KEEP STUFF SAFE

Introduction

- With unlimited $ easy to keep data forever
  - But $ are limited, so ...
- Inescapable tradeoff: $ vs. reliability
  - How do we make good tradeoff decisions?
- We need quantitative models of:
  - Reliability – these models are very hard:
  - How Few Copies? http://blog.dshr.org/2011/03/how-few-copies.html
  - Cost – these models are quite hard too
  - **NB – this is work in progress!**

A Lifetime of Decisions

- Keeping data safe is not a one-time thing
  - Life of data longer than life of hardware systems
    - Service lives vary through time
  - Systems, media have different purchase & running costs
    - Which vary though time
  - Interest rates to finance purchase vary through time
- At each stage of life of data, decision to take
  - Continue to use current technology?
  - Replace it with newer technology? If so, which?

DAWN

- Solid State Memory
  - Much more expensive to buy
  - Much cheaper to run, much longer service life
- Does it make sense to use it?
  - *Using Storage Class Memory for Archives with DAWN, a Durable Array of Wimpy Nodes*
  - http://www.ssrc.ucsc.edu/pub/adams-ssrctr-11-05.html
- We argue that it does make sense
  - On the basis of a simplistic economic analysis

Discounted Cash Flow

- Costs now vs. future costs?
  - Assume an interest rate
  - Invest now so that principal + interest = future cost
  - Amount invested now is net present value of future cost
- What interest rate to use?
  - If certain about future, use Treasury bond rates
  - The less certain, the more *risk premium* added to rate
- This is the standard technique investors use
  - What could possibly go wrong?

Does DCF Work In Practice?

First, there is statistically significant evidence of short-termism in the pricing of companies' equities. This is true across all industrial sectors. Moreover, there is evidence of short-termism having increased over the recent past. Myopia is mounting. Second, estimates of short-termism are economically as well as statistically significant. Empirical evidence points to excess discounting of between 5% and 10% per year.

- A. Haldane & R. Davies *The Short Long*

Does DCF Work In Theory?

What this analysis makes clear, however, is that the long term behavior of valuations depends extremely sensitively on the interest rate model. The fact that the present value of actions that affect the far future can shift from a few percent to infinity when we move from a constant interest rate to a geometric random walk calls seriously into question many well regarded analyses of the economic consequences of global warming. ... no fixed discount rate is really adequate – as our analysis makes abundantly clear, the proper discounting function is not an exponential.

- D. Farmer & J. Geankoplos *Hyperbolic Discounting is Rational: Valuing the Far Future with Uncertain Discount Rates*

Model

- A single interest rate is just wrong
  - So is a single rate of $/byte decrease
- Need to follow possible paths through future
  - Varying interest rates
  - Varying technology costs and service lives
- Need a Monte Carlo simulation
  - Follow many paths
  - Chose interest rates, costs, service lives from distributions
  - Average over the paths to get:
    - Most probable outcomes, range of outcomes

Components of the Model

- Yield Curves
  - Relate duration of loan to interest rate at given time
- Loans
  - Principal, interest rate, duration, $ outstanding
- Assets
  - Principal, interest rate, duration
- Technologies
  - Purchase cost, running cost

Technology

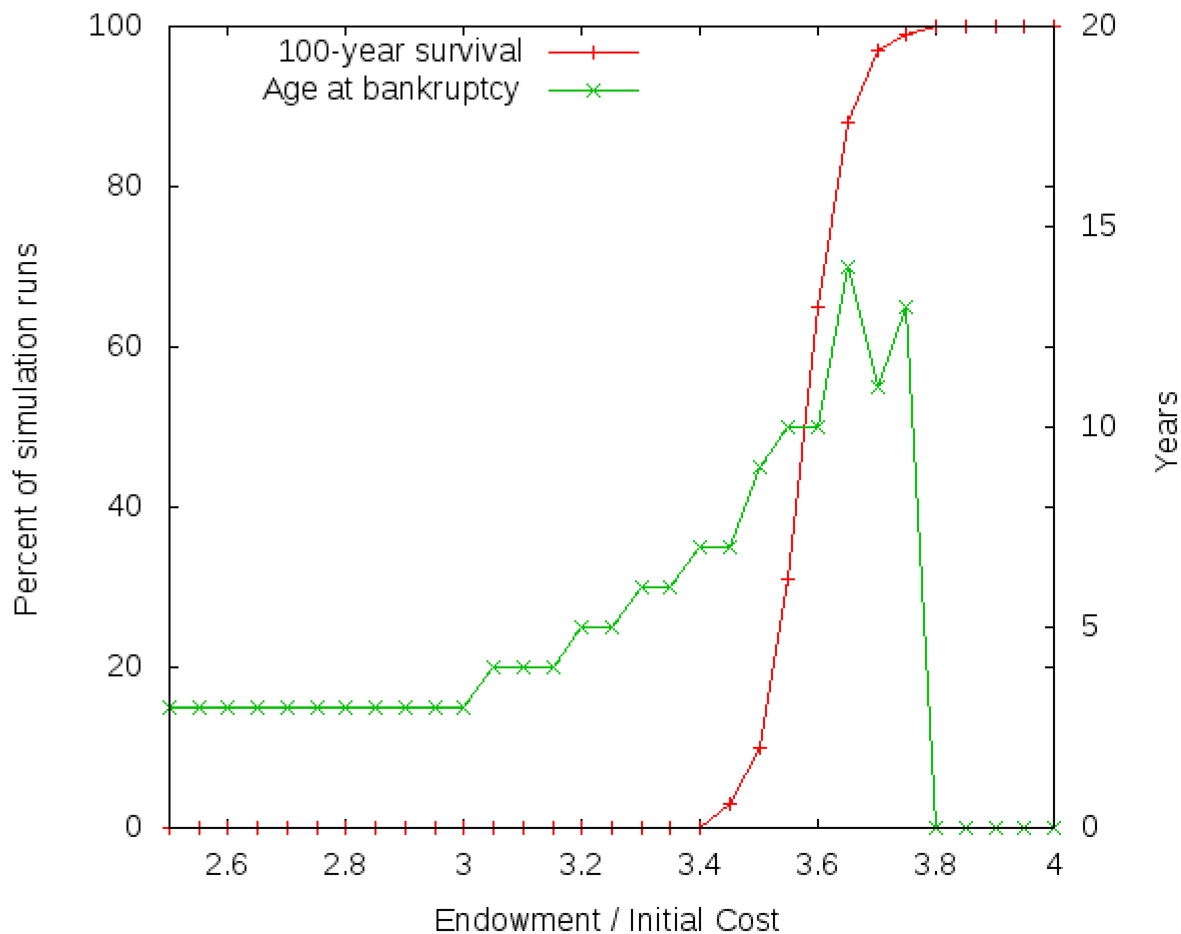
- Each technology has:
  - *Purchase cost*: e.g. for disk follows Kryder's Law
  - *Running cost*: e.g. for cloud, rental, bandwidth, compute
  - *Move-in cost*: e.g. for cloud upload bandwidth cost
  - *Move-out cost*: migration cost = move-out + move-in
  - *Service life*: e.g. for disk 4 or 5 years
- If model chooses to deploy a technology:
  - *Purchase loan*: pays for purchase and migration cost
  - Duration of loan = service life, interest from yield curve

Operations of the Model

- Choose a year at random – set interest rates
- Create some new technologies
  - If better, replace existing technologies in available set
- For each deployed technology:
  - If at end-of-life, replace by best available technology
  - Else, for each available technology:
    - Compute cost over shorter of remaining life & horizon
  - If cheapest available technology saves money, use it
  - Pay running cost and purchase loan cost
- Pay any outstanding purchase loans
  - From prematurely retired technologies

Example of Using the Model

- Kryder's Law
  - + running cost
- Interest rates
  - From 1990-2010
- Endow data
  - % survive?
  - How long to $0?



Other Uses of the Model

- Effects of short-termism on storage cost?
  - How long a planning horizon should we use?
- Effect on endowment of pauses in $/byte?
- Low running costs vs. low purchase costs?
  - E.g. flash vs. disk?
- Local disk vs cloud storage?
  - Purchase + run costs vs. run cost only?

Feedback?

- Does this model look useful to you?
- What other concepts does it need?
  - Replication policies
  - Data growth
  - ...
- What questions do you want to ask it?