On the Nature of Financial Risk: Why Risk is so Hard to Measure and Why Risk Models Fail so Often

Jón Daníelsson
Systemic Risk Centre
London School of Economics

www.ModelsAndRisk.org
www.SystemicRisk.ac.uk

New Frontiers in Systemic Risk Measures
June 4, 2015
“Model Risk of Risk Models”, (2014) with Kevin James (PRA), Marcela Valenzuela (University of Chile) and Ilknur Zer (Federal Reserve)

“Why risk is so hard to measure” (2015) with Chen Zhou, Bank of Netherlands and Erasmus University, 2015

And several VoxEU.org blogs
Some actual price series
Some actual price series (Zoom in)
Lets forecast risk...

with “reputable” models generally accepted by authorities and industry

- Value–at–Risk (VaR) and Expected Shortfall (ES)
- Probability 1%
- Using as model
  
  **MA** moving average  
  **EWMA** exponentially weighted moving average  
  **GARCH** normal innovations  
  **t–GARCH** student–t innovations  
  **HS** historical simulation  
  **EVT** extreme value theory

- Estimation period 1,000 days
## Risk for the next day \((t + 1)\)

Portfolio value is 1,000

<table>
<thead>
<tr>
<th>Model</th>
<th>VaR</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>14.04</td>
<td>20.33</td>
</tr>
<tr>
<td>MA</td>
<td>11.42</td>
<td>13.09</td>
</tr>
<tr>
<td>EWMA</td>
<td>1.59</td>
<td>1.82</td>
</tr>
<tr>
<td>GARCH</td>
<td>1.71</td>
<td>1.96</td>
</tr>
<tr>
<td>tGARCH</td>
<td>2.10</td>
<td>2.89</td>
</tr>
<tr>
<td>EVT</td>
<td>13.90</td>
<td>24.41</td>
</tr>
<tr>
<td>Model risk</td>
<td>8.85</td>
<td>13.43</td>
</tr>
</tbody>
</table>
Let's add one more day...
How frequently do the Swiss appreciate by 15.5%?
measured in once every $X$ years

Model frequency
How frequently do the Swiss appreciate by 15.5%?
measured in once every $X$ years

<table>
<thead>
<tr>
<th>Model</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>never</td>
</tr>
</tbody>
</table>
How frequently do the Swiss appreciate by 15.5%?
measured in once every $X$ years

<table>
<thead>
<tr>
<th>Model</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>never</td>
</tr>
<tr>
<td>GARCH</td>
<td>never</td>
</tr>
</tbody>
</table>
How frequently do the Swiss appreciate by 15.5%?

measured in once every $X$ years

<table>
<thead>
<tr>
<th>Model</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>never</td>
</tr>
<tr>
<td>GARCH</td>
<td>never</td>
</tr>
<tr>
<td>MA</td>
<td>$2.7 \times 10^{217}$</td>
</tr>
</tbody>
</table>
How frequently do the Swiss appreciate by 15.5%?
measured in once every \(X\) years

<table>
<thead>
<tr>
<th>Model</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>never</td>
</tr>
<tr>
<td>GARCH</td>
<td>never</td>
</tr>
<tr>
<td>MA</td>
<td>(2.7 \times 10^{217}) age of the universe is about (1.4 \times 10^{10})</td>
</tr>
<tr>
<td>tGARCH</td>
<td>(1.4 \times 10^{7}) age of the earth is about (4.5 \times 10^{9})</td>
</tr>
</tbody>
</table>
### How frequently do the Swiss appreciate by 15.5%?

**measured in once every $X$ years**

<table>
<thead>
<tr>
<th>Model</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>never</td>
</tr>
<tr>
<td>GARCH</td>
<td>never</td>
</tr>
<tr>
<td>MA</td>
<td>$2.7 \times 10^{217}$</td>
</tr>
<tr>
<td>tGARCH</td>
<td>$1.4 \times 10^7$</td>
</tr>
<tr>
<td>EVT</td>
<td>109</td>
</tr>
</tbody>
</table>

*Age of the universe is about $1.4 \times 10^{10}$
Age of the earth is about $4.5 \times 10^9$*
How frequently do the Swiss appreciate by \(15.5\%\) measured in once every \(X\) years

<table>
<thead>
<tr>
<th>Model</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>never</td>
</tr>
<tr>
<td>GARCH</td>
<td>never</td>
</tr>
<tr>
<td>MA</td>
<td>(2.7 \times 10^{217})</td>
</tr>
</tbody>
</table>
| tGARCH  | \(1.4 \times 10^7\) | age of the universe is about \(1.4 \times 10^{10}\)  
age of the earth is about \(4.5 \times 10^9\) |
| EVT     | 109       |
Even more interesting after the event
Even more interesting after the event
Even more interesting after the event

![Chart showing HS, MA, and EVT performance over time]

- HS
- MA
- EVT

Jan 01  Jan 15  Feb 01  Feb 15

−15%  −10%  −5%  0%
Even more interesting after the event

Jan 01 Jan 15 Feb 01 Feb 15

MA

EWMA

GARCH

HS

EVT

-15%

-10%

-5%

0%
Even more interesting after the event
So

- Depending on model risk, may or may not, not move
- Some models signal very high risk when we know nothing else will happen
- Can go to www.ModelsAndRisk.org/forecast/ for more details and more assets
But is the event all that extraordinary?

just eyeballing it seems not that much
Could we do better?

- If one considers who owns the Swiss National Bank
- And some factors, perhaps
  - SNB dividend payments
  - Money supply
  - Reserves
  - Government bonds outstanding
- Yes, we can do much much better than the models used here
- But they are what is prescribed

example is from http://www.voxeu.org/article/what-swiss-fx-shock-says-about-risk-models
Systemic Risk
“Systemic crises” in OECD countries from 1970 – 2011
from the IMF systemic crises database

<table>
<thead>
<tr>
<th>crises</th>
<th>#</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>Australia, Canada, New Zealand</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Netherlands, Norway, Poland, Portugal, Slovak Republic, Switzerland,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>United Kingdom</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>Chile, Hungary, Mexico, Slovenia, Spain, Sweden, Turkey, United States</td>
</tr>
</tbody>
</table>
Summary stats

- Weighted by year of joining OECD
- Unconditional probability of a crisis is 2.33% a year, or once every 43 years
- This suggests that target probabilities should be set to
  - $p = 0.023$ per year
  - $p = 9 \times 10^{-5}$ per day
Many — even most — statistical models aiming to help us to understand systemic risk

Make stronger assumptions about how history repeats itself in a statistical sense

but the empirical evidence suggests this is not true
• Very little statistical regularity in crises
  • Volatility may fall or increase
  • Tails may be thicker or thinner
  • Etc.
• Statistical modeling of crises does not seem to deliver very much
• A logical conclusion of the earlier endogenous risk analysis
Classifying systemic risk measures

Danielsson et. (2014) “Model risk of risk models”

<table>
<thead>
<tr>
<th>Marginal risk measure</th>
<th>Condition on system</th>
<th>Condition on institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>VaR</td>
<td>pr[(R_i \leq Q_i</td>
<td>R_s \leq Q_s)] = p</td>
</tr>
<tr>
<td>MVaR</td>
<td></td>
<td>CoVaR</td>
</tr>
<tr>
<td>MES</td>
<td>E[(R_i</td>
<td>R_s \leq Q_s)]</td>
</tr>
<tr>
<td>ES</td>
<td></td>
<td>E[(R_s</td>
</tr>
</tbody>
</table>

- SES/MES/CoVaR/SRISK/* all are elementally based on VaR
- None remotely captures systemic risk
- Systemic crisis happen *every 42 years* on average, not *3–12 times a year* as implied those
- It is very hard mathematically to translate risk across probability levels
“Model Risk of Risk Models”

(2014)

with Kevin James (PRA)
Marcela Valenzuela (University of Chile)
Ilknur Zer (Federal Reserve)
<table>
<thead>
<tr>
<th>Case study</th>
<th>Systemic risk</th>
<th>Model Risk</th>
<th>Myth and reality</th>
<th>Nature of risk</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

**EUROMONEY**

Volume 40, Number 485

September 2009

**Banking's Next Top Model**
Model risk of risk forecast models

Every model is wrong — Some models are useful

The risk of loss, or other undesirable outcomes like financial crises arising from using risk models to make financial decisions

- Infinite number of candidate models
- Infinite number of different risk forecasts for the same event
- Infinite number of different decisions, many ex ante equally plausible
- Hard to discriminate
Do we care?

- Much *anecdotal grumbling*
- The common wisdom maintains that models failed to cover themselves in glory before 2007
- The models today are not much different from the models then
- *So*
- Why are they becoming more and more common
- Why is there so little scrutiny of them (beyond grumbling and tick the box exercises)?
Risk ratios

our proposed model risk methodology

- Consider the problem of forecasting risk for day $t + 1$ using information available on day $t$
- Suppose we have $N$ candidate models to forecast the risk, each providing different forecasts

\[ \text{RR}_{t+1} = \frac{\max \left\{ \text{Risk}^n_{t+1} \right\}_{n=1}^N}{\min \left\{ \text{Risk}^n_{t+1} \right\}_{n=1}^N} \]

- We then define model risk as the ratio the highest to the lowest risk forecasts
Model choice

- **MA** moving average
- **EWMA** exponentially weighted moving average
- **GARCH** normal innovations
- **t–GARCH** student–t innovations
- **HS** historical simulation
- **EVT** extreme value theory

- All models re–estimated every day

We can, and have, tried the new shiny.
Each new model will weakly increase the RR
Risk measures and data

- Current Basel: VaR 99%
- Proposed Basel III: ES 97.5%, overlapping estimation windows
- Large financials traded on the NYSE, AMEX, and NASDAQ
  - banking, insurance, real estate, and trading sectors
- January 1970 to December 2012.
- Sampling frequencies daily
- Sample size shown here 1,000 days
## Sample results

**JPM January 3, 2007, $100 portfolio**

<table>
<thead>
<tr>
<th>Model</th>
<th>VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>$3.22</td>
</tr>
<tr>
<td>MA</td>
<td>$2.91</td>
</tr>
<tr>
<td>EWMA</td>
<td>$1.96</td>
</tr>
<tr>
<td>GARCH</td>
<td>$2.13</td>
</tr>
<tr>
<td>tGARCH</td>
<td>$2.74</td>
</tr>
<tr>
<td>EVT</td>
<td>$3.22</td>
</tr>
</tbody>
</table>

**Model risk** 1.64
JPM

Model risk (risk ratios)
Model risk — Across all assets

Annual maximum

- mean
- 95% conf interval

<table>
<thead>
<tr>
<th>Year</th>
<th>1980</th>
<th>1990</th>
<th>2000</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% conf interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Why risk is so hard to measure”, 2015

with Chen Zhou, Bank of Netherlands and Erasmus University, 2015
Objective

- What is the relationship between ES and VaR?
  - VaR(99%) and ES(97.5%) because of Basel
- What are the small sample properties of these risk measures?
- What is the implication of using overlapping estimation windows?
- Risk measures compared by Monte Carlo simulations
  - $10^7$ simulations (yes, we need that many)
- And theoretic analysis
- Across sample sizes and tail thicknesses
ES(99%)/VaR(99%)

- Theory
- N = 2 years
- N = 200 days
- N = 50 years

ES/VaR vs. tail thickness (smaller is thicker)
ES(97.5%)/VaR(99%)
Finite sample properties of VaR

\[ \alpha = 3 \]

- **true VaR**
- **VaR estimate**
- **99% confidence interval**
Finite sample properties of VaR

\( N = 1000 \)

- **true**
- **Q99**
- **mean**

<table>
<thead>
<tr>
<th>Tail thickness (smaller is thicker)</th>
<th>VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td></td>
</tr>
</tbody>
</table>
Finite sample properties of VaR

\[ N = 300 \]

- **true**
- **Q99**
- **mean**
Is ES really better than VaR?

yes, I know it is subadditive

- VaR is also subadditive unless tails are superfat
  - (tail index $< 2$)
- In practice, ES is VaR times a constant
  - Affected by tail thickness and sample size
- ES is less precisely estimated than VaR
- With the distributions and probabilities considered here, VaR is preferred to ES
Estimation with overlapping data

- 10 and 50 day windows
- Also t–GARCH and CRSP data
- Compare ($N$ sample size, $H$ overlap interval)
  1. $H + N$–day overlapping estimation
  2. $N$ days with $\sqrt{H}$ scaling
- $\sqrt{H}$ scaling is more robust than estimation with overlapping data
Conclusion

- VaR beats ES
  - Only reason to prefer ES is when concerned with manipulation
- Overlapping estimation cannot be recommended
- Minimum sample size thousand days
Nature of risk
Why models perform the way they perform

1. The statistical theory of the models
2. The nature of risk
Forecasting a tail when we know the distribution

- Asymptotically everything might be fine but what are the *small sample properties*?
- With a properly specified model, a 99% confidence interval may be (VaR=1)
  - 500 observations
    \[ \text{VaR} = \text{runif}() \]
  - 1,000 observations,
    \[ \text{VaR} \in [0.7, 1.6] \]
  - 10,000 observations
    \[ \text{VaR} \in [0.9, 1.13] \]
And in the real world

- Where returns follow an unknown stochastic process
- The uncertainty about the risk forecasts will be much higher
- This goes a long way to explain why different risk models, each plausible, can give such widely differing results
The nature of risk

- We have classified risk as *exogenous* or *endogenous*

  **exogenous** Shocks to the financial system arrive from outside the system, like with an asteroid

  **endogenous** Financial risk is created by the interaction of market participants

“The received wisdom is that risk increases in recessions and falls in booms. In contrast, it may be more helpful to think of risk as increasing during upswings, as financial imbalances build up, and materialising in recessions.”

Andrew Crockett, then head of the BIS, 2000
Risk is endogenous

- Market participants are guided by a myriad of models and rules, many dictate myopia
- Prices are not Markovian

Risk models underestimate risk during calm times and overestimate risk during crisis — they get it wrong in all states of the world
Two faces of risk

- When individuals observe and react — affecting their operating environment
- Financial system is not invariant under observation
- We cycle between virtuous and vicious feedbacks
  - risk reported by most risk forecast models — perceived risk
  - actual risk that is hidden but ever present
Endogenous bubble

![Graph showing an endogenous bubble with Prices on the y-axis and a peak at t=13]

Prices
Endogenous bubble

- Prices
- Perceived risk

Diagram showing the relationship between prices and perceived risk over time.
Endogenous bubble

- Prices
- Perceived risk
- Actual risk
The lessons are...

- Risk is created out of sight in a way that is not detectable
- Attempts to measure risk — especially extreme risk — are likely to fail
  - systemic risk measures like CoVaR, SES/MES, Sharpley, SRisk do not remotely capture systemic risk
  - every systemic risk estimation method that is based on market data is likely to fail
- Why?
  - endogenous risk
  - stability is destabilizing
  - market prices react after event happens
Conclusion
It matters what models are used for and how they are used

- Risk models are
  
  **most useful for** risk controlling traders
  **less useful in** risk capital allocation
  **mostly useless for** financial regulations
  **dangerous when used for** macro–prudential policy

one better not fall into the trap of doing probability shifting
Harmonization

- If we regulate by models we must believe there is one true model
- Therefore, banks should not report different risk readings for the same portfolio
- However, forcing model harmonization across banks is pro–cyclical
- And forcing the same models to be used for everything internally is also pro–cyclical
- And pro–cyclicality negatively affects economic growth and increases financial instability

model harmonization cannot be recommended for macro–prudential reasons
So

- Risk models are subject to considerable model risk, but the signal is often useful
- If one understands the model risk of risk models, they can provide a useful guidance
- Concern that important policy decisions are based on such poor numbers
- Basic compliance suggests that risk models outcomes should contain *confidence bounds*
The cost of a type I or type II error is significant.

The minimum acceptable criteria for a risk model should not be to weakly beat noise.