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.SystemicRisk.ac.uk

February 24, 2016
SYSTEMIC RISK: WHAT RESEARCH TELLS US AND WHAT WE NEED TO FIND OUT
The presentation is based on

- “Model Risk of Risk Models”, (2016) with Kevin James (PCA and LSE), Marcela Valenzuela (University of Chile) and Ilknur Zer (Federal Reserve), forthcoming Journal of Financial Stability
- “Why risk is so hard to measure” (2016) with Chen Zhou, Bank of Netherlands and Erasmus University, 2015
- “Learning from History: Volatility and Financial Crises” (2016) with Marcela Valenzuela (University of Chile) and Ilknur Zer (Federal Reserve)
- 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><strong>Model risk</strong></td>
<td><strong>8.85</strong></td>
<td><strong>13.43</strong></td>
</tr>
</tbody>
</table>
Let's add one more day...
€/CHF

[Graph showing the exchange rate of €/CHF from 2000 to 2015 with significant drops in 2009 and 2011.]
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>

Age of the universe is about $1.4 \times 10^{10}$
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>
</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>
<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>
<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 $4.5 \times 10^9$
Even more interesting after the event
Even more interesting after the event

![Graph showing HS and EVT trends from Jan 01 to Feb 15]

- Jan 01
- Jan 15
- Feb 01
- Feb 15
Even more interesting after the event

![Graph showing financial performance with lines labeled HS, MA, and EVT. The x-axis represents dates from January 01 to February 15, with a significant drop in performance around January 15.]
Even more interesting after the event
Even more interesting after the event
### Case Study

- 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/](http://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 www.voxeu.org/article/what-swiss-fx-shock-says-about-risk-models
“Model Risk of Risk Models”

(2016)

with Kevin James (PRA)
Marcela Valenzuela (University of Chile)
Ilknur Zer (Federal Reserve), forthcoming Journal of Financial Stability
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

$$\{ \text{Risk}^n_{t+1} \}_{n=1}^N$$

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

$$\text{Risk Ratio}_{t+1} = RR_{t+1} = \frac{\max \{ \text{Risk}^n_{t+1} \}_{n=1}^N}{\min \{ \text{Risk}^n_{t+1} \}_{n=1}^N}$$
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)
Zooming in (end of quarter)

VaRs

2007 2008 2009 2010
Zooming in (end of quarter)

VaRs
Model risk — Across all assets

Annual maximum
Model risk — Across all assets

Annual maximum
Model risk — Across all assets

Annual maximum

- mean
- 95% conf interval
Model risk and market conditions I

- Graphs suggest model risk is typically quite moderate, but volatile and sharply increases during some periods.
- Any relationship with the overall market risk?
Model risk and market conditions I

- Graphs suggest model risk is typically quite moderate, but volatile and sharply increases during some periods.
- Any relationship with the overall market risk?
- Compare the model risk with the VIX
  - Significant correlation with model risk ($\rho = 19.2\%$)
  - Model risk does not Granger cause VIX
  - The VIX Granger causes model risk
Similar models produce similar forecasts

Average across time and stocks

<table>
<thead>
<tr>
<th></th>
<th>EWMA</th>
<th>GARCH</th>
<th>tGARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>1</td>
<td>0.874</td>
<td>0.848</td>
</tr>
<tr>
<td>GARCH</td>
<td></td>
<td>1</td>
<td>0.905</td>
</tr>
<tr>
<td>tGARCH</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

- Risk readings of GARCH models are highly correlated
**Similar models produce similar forecasts**

Average across time and stocks

<table>
<thead>
<tr>
<th></th>
<th>EWMA</th>
<th>GARCH</th>
<th>tGARCH</th>
<th>HS</th>
<th>MA</th>
<th>EVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>1</td>
<td>0.874</td>
<td>0.848</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td></td>
<td>1</td>
<td>0.905</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tGARCH</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.902</td>
<td>0.964</td>
</tr>
<tr>
<td>MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.934</td>
</tr>
<tr>
<td>EVT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

- Risk readings of GARCH models are highly correlated
- Readings of HS, MA, and EVT are highly correlated
Similar models produce similar forecasts
Average across time and stocks

<table>
<thead>
<tr>
<th></th>
<th>EWMA</th>
<th>GARCH</th>
<th>tGARCH</th>
<th>HS</th>
<th>MA</th>
<th>EVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWMA</td>
<td>1</td>
<td>0.874</td>
<td>0.848</td>
<td>0.258</td>
<td>0.248</td>
<td>0.267</td>
</tr>
<tr>
<td>GARCH</td>
<td></td>
<td>1</td>
<td>0.905</td>
<td>0.310</td>
<td>0.309</td>
<td>0.320</td>
</tr>
<tr>
<td>tGARCH</td>
<td></td>
<td></td>
<td>1</td>
<td>0.308</td>
<td>0.300</td>
<td>0.317</td>
</tr>
<tr>
<td>HS</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.902</td>
<td>0.964</td>
</tr>
<tr>
<td>MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.934</td>
</tr>
<tr>
<td>EVT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

- Risk readings of GARCH models are highly correlated
- Readings of HS, MA, and EVT are highly correlated
- The inconsistency is between the two “groups”
Each model has its own merits

There is no clear conclusion of which one is better

Particular models perform best in a particular environments and purposes

- React quickly to news—volatility-based, like GARCH
- Easy to compute—EWMA, MA, HS
- Very small samples—EWMA, MA
- Low volatility of risk —HS
- Tails —EVT, tGARCH
Big picture

• As expected similar models create similar risk forecasts
• However even in a given model, parameters may significantly affect the risk readings
• Including new models to the analysis can only increase the disagreement
• To identify the best model is *not* the aim of this paper
• The aim is to underline: Concluding Bank X (uses HS) being riskier than Bank Y (uses GARCH) requires caution
“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
Simulation schedule

- Student-t (2,3,4,5,6)
- Pareto (2,3,4,5,6)
- Student-t GARCH
- $N = 300, 1000, 2500, 12500$
- Simulations $10^7$
Accuracy of VaR (ES)

- We know the asymptotic properties
- But what happens when the sample size becomes smaller and smaller
- Across various tail thicknesses
Finite sample properties of VaR

\( \alpha = 3 \)

- **true VaR**
- **VaR estimate**
- **99% confidence interval**

<table>
<thead>
<tr>
<th>Years</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Finite sample properties of VaR

\( N = 1000 \)

- Black line: true
- Blue line: Q99
- Red line: mean

Tail thickness (smaller is thicker) vs. VaR
Finite sample properties of VaR

\[ N = 300 \]

Tail thickness (smaller is thicker)

- **true**
- **Q99**
- **mean**
ES is increasingly popular (ask the Basel Committee and Solvency II people)

Certainly, provided tail index $> 1$ it is theoretically superior to VaR

But what about its small sample properties?

It has higher estimation uncertainty than VaR

And different bias

So, what about the ratio of ES to VaR

At the 99% and the Basel III 97.5%
ES(99%)/VaR(99%)

- **Theory**
- **N = 2 years**
- **N = 200 days**
- **N = 50 years**
ES(97.5%)/VaR(99%)
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, preferably more
- At lower sample sizes, might as well use a random number generator
<table>
<thead>
<tr>
<th>Case study</th>
<th>Model risk</th>
<th>Measure risk</th>
<th>Nature of risk</th>
<th>Minsky</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Nature of risk</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Prices
Endogenous bubble

- Prices
- Perceived risk
Endogenous bubble

- Prices
- Perceived risk
- Actual risk
Risk, macro pru and stress testing

- Systemic crisis happen every 42 years (every 17 in UK)
- A risk forecast should be tailored to that event frequency
- One should recognize that actual risk builds up out of sight in a way that is not detectable
- Best to minimize the use of risk forecasting as much as possible
The 42 year cycle of systemic risk

actual risk builds up

2000 2010 2020 2030 2040
The 42 year cycle of systemic risk

(actual risk builds up)

2000 2010 2020 2030 2040
The 42 year cycle of systemic risk

perceived risk indicators flash

actual risk builds up

hidden trigger

2000 2010 2020 2030 2040
The 42 year cycle of systemic risk

- perceived risk indicators flash
- actual risk builds up
- improvised responses
- hidden trigger

2000 2010 2020 2030 2040
The 42 year cycle of systemic risk

- The 42 year cycle of systemic risk.
- Actual risk builds up.
- Perceived risk indicators flash.
- Hidden trigger.
- Improvised responses.
- MacroPru implemented.
- 2000 to 2040 timeline.
The 42 year cycle of systemic risk

- actual risk builds up
- perceived risk indicators flash
- improvised responses
- MacroPru implemented
- actual risk builds up

2000 → 2010 → 2020 → 2030 → 2040
The 42 year cycle of systemic risk
The 42 year cycle of systemic risk

Perceived risk
indicators flash

improvised responses

hidden trigger

MacroPru implemented

2000 2010 2020 2030 2040
The 42 year cycle of systemic risk

- Perceived risk
  - indicators flash
- Improvised responses
- Hidden trigger
  - MacroPru implemented

The 42 year cycle
“Learning from History: Volatility and Financial Crises”

(2015)

with Marcela Valenzuela (University of Chile)
Ilknur Zer (Federal Reserve)
Minsky

- Economic agents perceive a low risk environment as a signal to increase risk-taking
- Which eventually leads to a crisis

“Stability is destabilizing”

“Volatility in markets is at low levels, both actual and expected, ... to the extent that low levels of volatility may induce risk-taking behavior ... is a concern to me and to the Committee.”

Federal Reserve Chair Janet Yellen, 2014.
Learning from History: Volatility and Financial Crises

- No extant empirical literature documenting such a relationship between financial market volatility, the real economy and crises
- We construct a comprehensive database on historical volatilities from primary sources (1800 to 2010, 60 countries)
- Volatility *does not* predict crises
- but
• Decomposing volatility into unexpectedly low and high volatilities
• Strong and significant relationship between unexpected volatilities and the likelihood of financial crises
• Unexpectedly low volatility increases the probability of both banking and stock market crises
• Especially strong if low volatility persists half a decade or longer.
• Low volatility significantly increases risk-taking (credit-to-GDP)
• For stock market crises, but not banking crises, high volatility also increases the likelihood of a crisis, but only with much shorter lags, up to two or three years.
Conclusion
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
It matters what models are used for and how they are used

- Risk models are
  - most useful for risk controlling traders
  - less useful in internal risk capital allocation
    1. e.g. invest in European equities or JPG
  - often useless for financial regulations
    1. Traders read things like Basel III as manual for where to take risk
  - 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* (EBA now discussing some)
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.