Measuring Market Risk with Entropy

Martin Vesely

Czech National Bank, Risk Management

Modern Tools for Financial Analysis and Modeling

Bratislava

30 May 2019

Content

- What is entropy?
- Why use entropy for market risk evaluation?
- Mathematical model
- How to calculate entropy in practice?
- Results
  - US share markets – "history of crises"
  - EUR/CZK rate and CNB rate commitment
  - Actual CNB portfolios
  - Entropy and behavioural finance
  - Conclusion

Simply speaking, entropy is measure of disorder, uncertainty or surprise.

Image source: pixabay.com

What is entropy? (2)

Branch	Measure of…	Low entropy	High entropy
Thermodynamics	…particle disorder		
Information Theory	…message uncertainty or surprise	*“Sun will rise tomorrow”*	*“It will be sunny or rainy tomorrow”*
Risk Management	….market volatility or surprising P&L outcome		

Images source: pixabay.com

Motivation: Why entropy?

1. No assumptions about underlying distribution
2. Portfolio diversification leads to decrease in entropy by definition – subadditivity (*not true for VaR*)
3. More robust than standard deviation
4. Capped for distributions on finite interval (*not true for standard deviation*)
5. Always exists *(not true for standard deviation, e.g. Cauchy or other fat-tail distributions)*
6. Easy to interpret as measure of surprise

Mathematical model of entropy (1)

- Consider two possible outcomes of experiment with probabilities *p* and *1-p*

- If p = 0 or p = 1 there is no uncertainty:

  only one outcome is possible and it always occurs

- However for p = ½ uncertainty is at maximum:

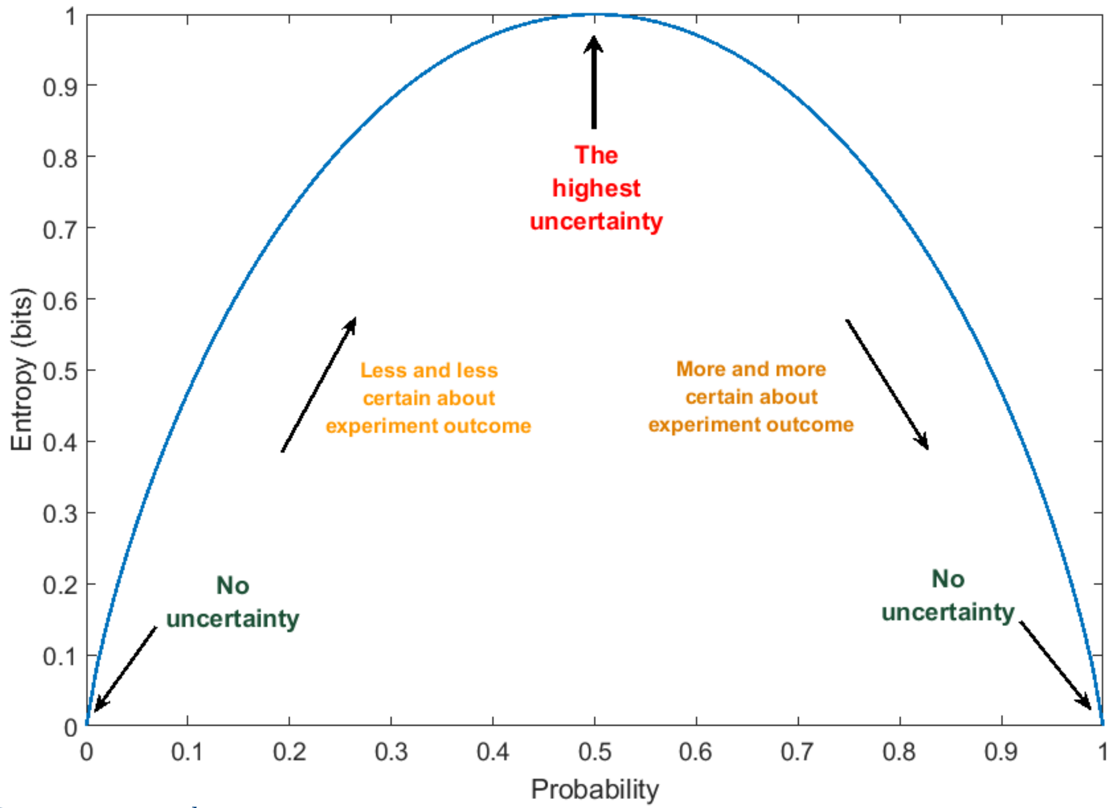
  50:50 $\Rightarrow$ no idea which outcome more likely to occur

- We are looking for function f(*p*) fulfilling
  - f(0) = f(1) = 0
  - Maximum occurs for p = 0.5

Mathematical model of entropy (2)

- One possible function is Shannon entropy:

$$H = -[p\log_a(p) + (1-p)\log_a(1-p)]$$



Source: own work

Notes:

1) *a = 2 for graph ⇒ entropy is expressed in bits (units will be discussed later)*
2) *0 log 0 is defined as 0 (i.e. value of p log(p) limit for p approaching zero)*

Mathematical model of entropy (3)

- Shannon entropy can be generalized for:

1. Discrete distributions (incl. distributions with $n \to \infty$)

$$H = -\sum_{i=1}^{n} p_i \log_a(p_i)$$

2. Continuous distributions ("reduced entropy")

$$H = -\int_R f(x) \log_a [f(x)] dx$$

Note: Reduced entropy can be negative. Lowest value (i.e. "no uncertainty") is $-\infty$.

Entropy and perfect certainty (1)

- Perfect certainty $\Rightarrow$ obviously zero entropy

- Inspired by 3rd law of thermodynamics:

  "Entropy of every system at absolute zero can be taken to be equal to zero"

- Shannon entropy of discrete distribution is zero by definition, but this is not the case for reduced entropy

- Total entropy of continuous distribution is in fact given by

$$H = -\int_R f(x) \log_a [f(x)] dx - \log_a [\Delta x]$$

Entropy and perfect certainty (2)

- Term $-\log_a[\Delta x]$ is "residuum" of switching from discrete to continuous realm
- Term approaches infinity as "delta" becomes "dee"
- To have H = 0, integral has to approach minus infinity for perfect certainty
  (!intuitive explanation, not mathematically fully correct!)
- **Reduced entropy** should be **used only for**:
  - Peer comparison
  - Time comparison
- **Overall, sufficient for use in finance**

Entropy units

- Entropy unit name depends on logarithm base used
  - $a = 2 \Rightarrow$ **bit**
  - $a = 3 \Rightarrow$ **trit**
  - $a = 10 \Rightarrow$ **dit**
  - $a = e\ (\approx 2.71) \Rightarrow$ **nat**
- Bits and nats are often used because…
  - …bits are usual measure of information content (*entropy of x bit means that one has to use x binary numbers for message encoding*)
  - …nats are connected with natural logarithm
- Entropy can be converted from base *a* to base *b* by dividing by $\log_a b$

How to calculate entropy?

1. Histogram estimator:
   - Based on definition for discrete case

$$\hat{H} = -\sum_{i=1}^{k} \frac{n_i}{n} \log_a\left(\frac{n_i}{n}\right) + \log_a(h)$$

Correction to bin width

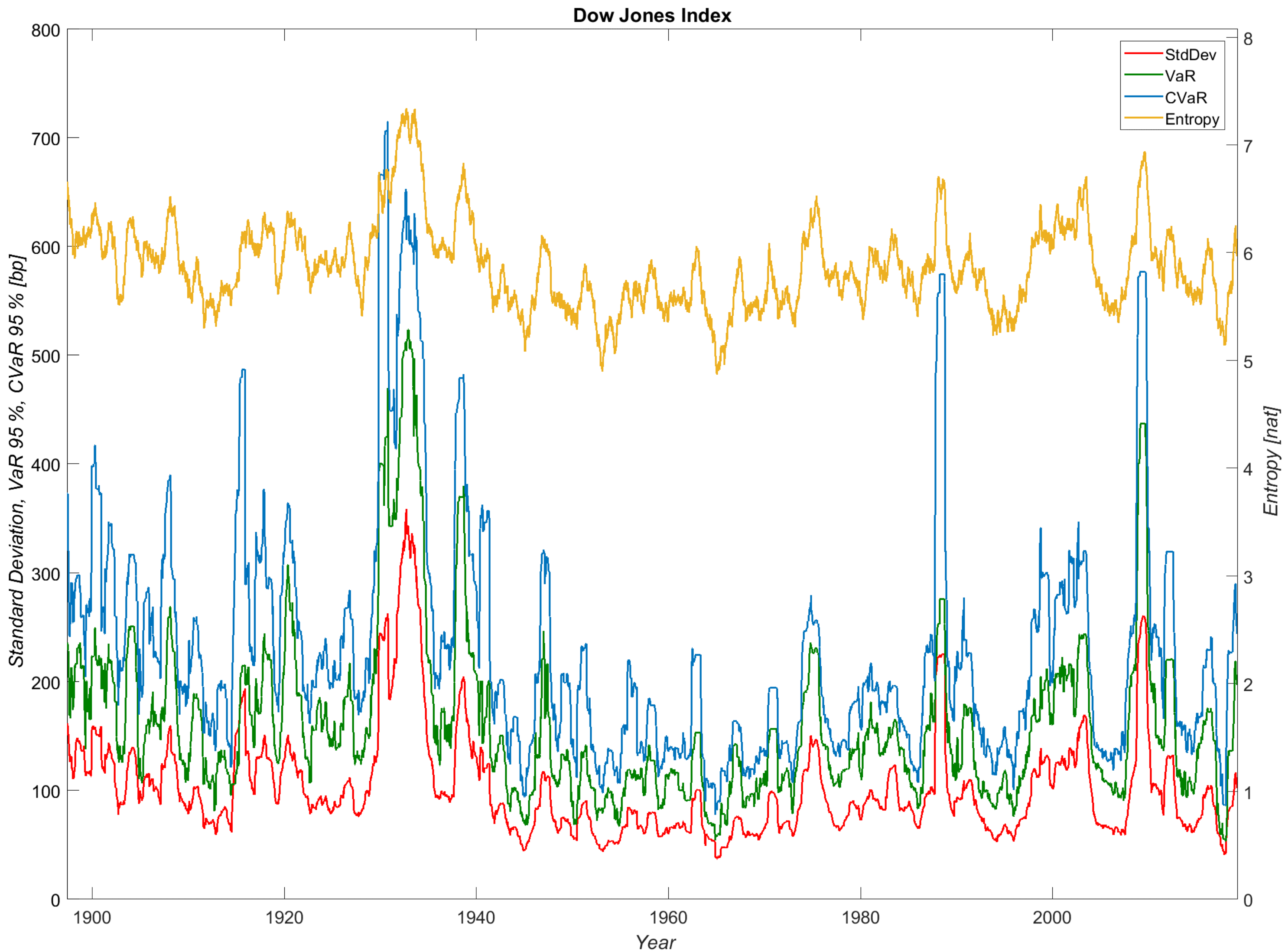
   - h…width of histogram bins, k…number of bins
     n_i…# observations in i^{th} bin, n…total # of observations
2. Kozachenko-Leonenko estimator (1D data):

$$\hat{H} = \frac{1}{n}\sum_{i=1}^{n} \log_a(r_i) + \log_a[2(n-1)] + \gamma$$

   - r_i is distance of observation to its nearest neighbour:
     $r_i = \min\{a_{i+1} - a_i; a_i - a_{i-1}\}$ for sorted observations a_i
   - If $r_i = 0$ then $r_i := 1/\sqrt{n}$
   - $\gamma \approx 0.5772156649$ (Euler-Mascheroni constant)

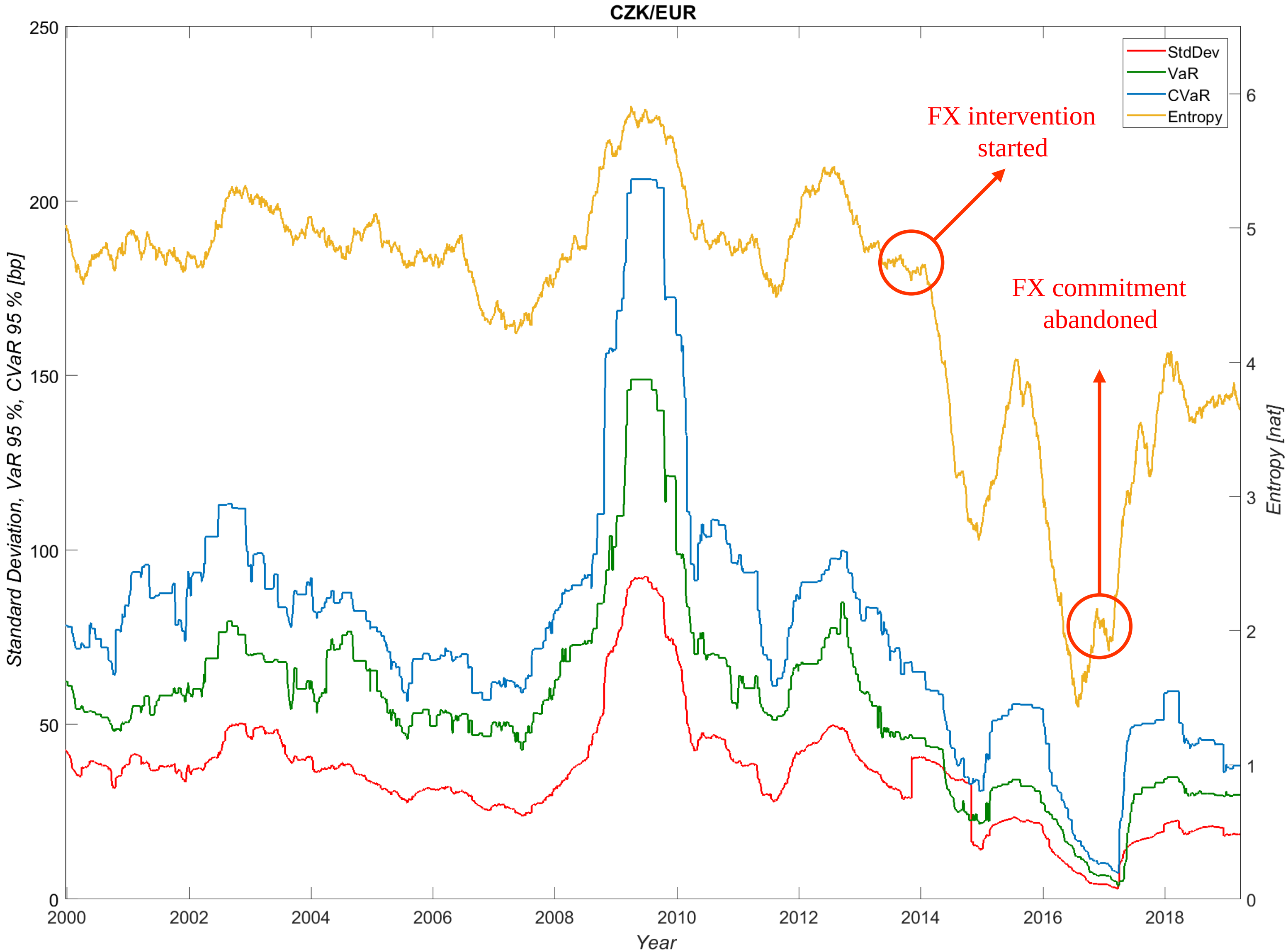
- Note: observation is actual value of e.g. P/L

Results – US equity markets



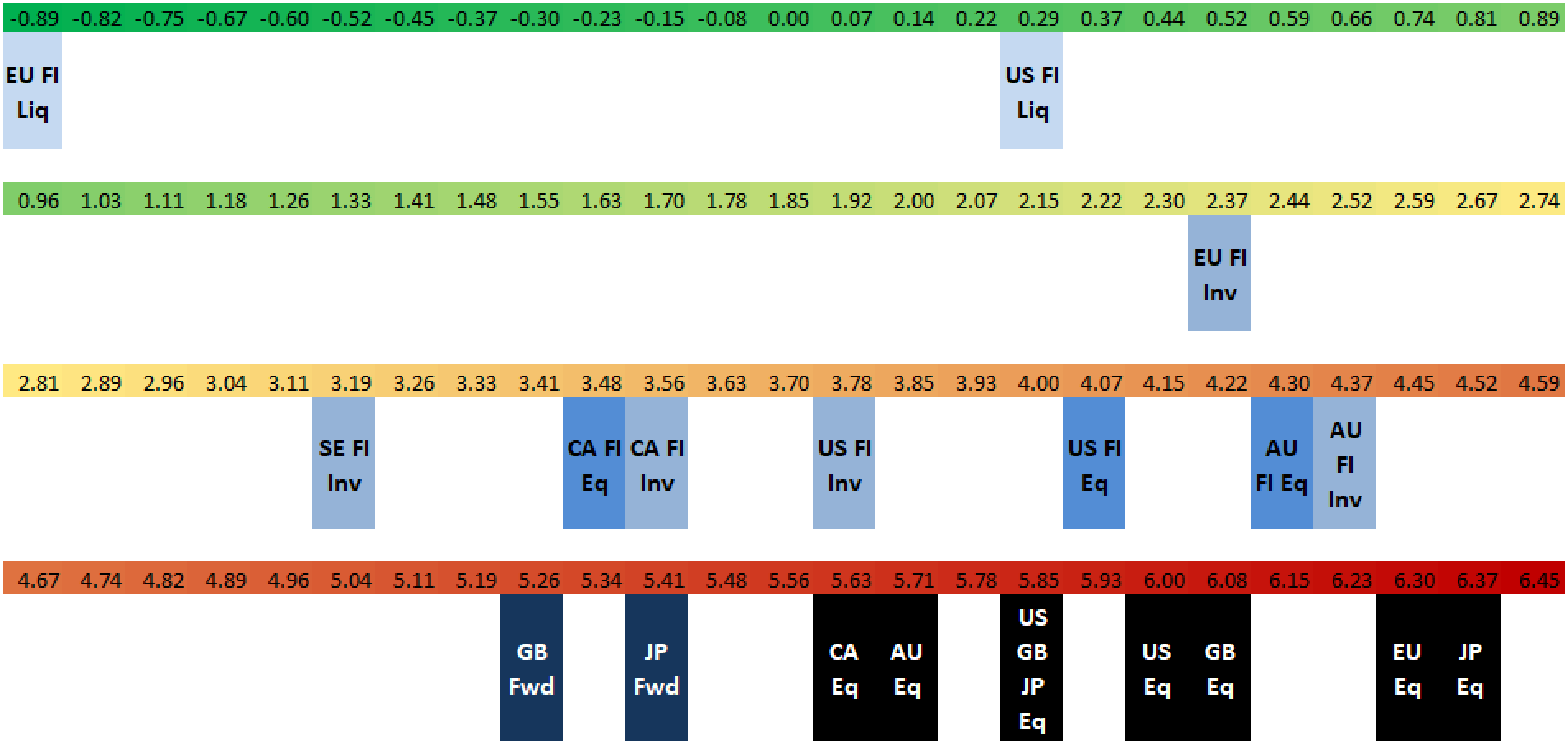
Source: Bloomberg, own calculation

Results – EUR/CZK



Source: Bloomberg, own calculation

Results – actual portfolios (03/2019)



Source: own calculation

legend:

FI…Fixed Income *Eq…Equity (shares)*

Liq…Liquidity ptf *Inv…Investment ptf*

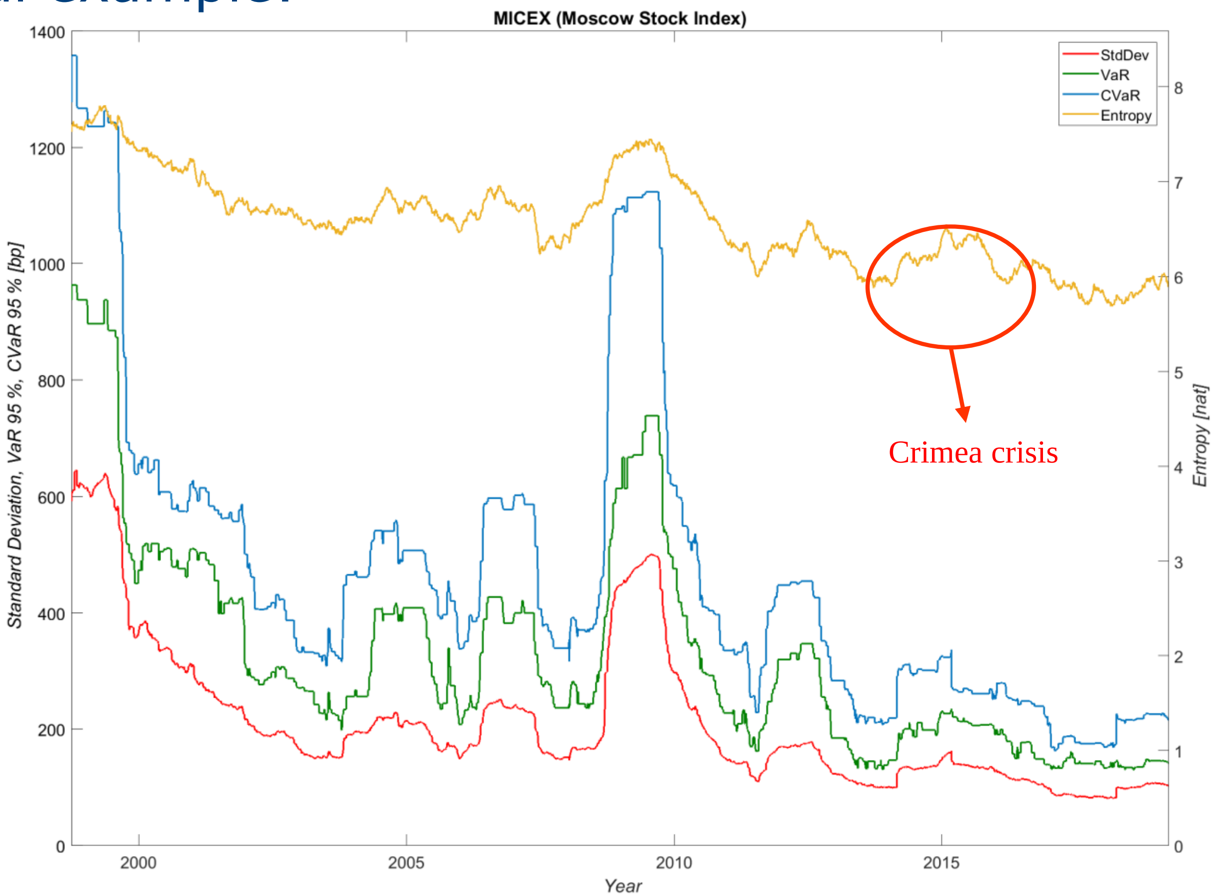
Fwd…Artificial cash position (currency overlay)

Results – entropy and behavioural finance (1)

- Entropy is more robust than standard deviation
  - less susceptible to outliers
  - $\Rightarrow$ less "overreaction" of market
- Theoretical example:
  - 10,000 random numbers distributed according to N(0,1) $\Rightarrow$ std. dev. is 0.979, entropy 1.387
  - Add 8 outliers: -7, -4.5, -4.5, -3, 3.9, 5, 5, 5.1
  - $\Rightarrow$ std. dev. is 0.987 (increase 0.8%), entropy 1.393 (increase 0.47%)
- Since entropy does not change as rapidly as standard deviation, investors should not overreact
- As a result, markets could be calmer, with only "shallow" crises

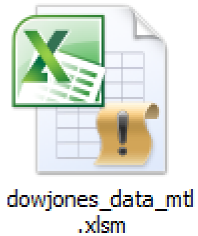
Results – entropy and behavioural finance (2)

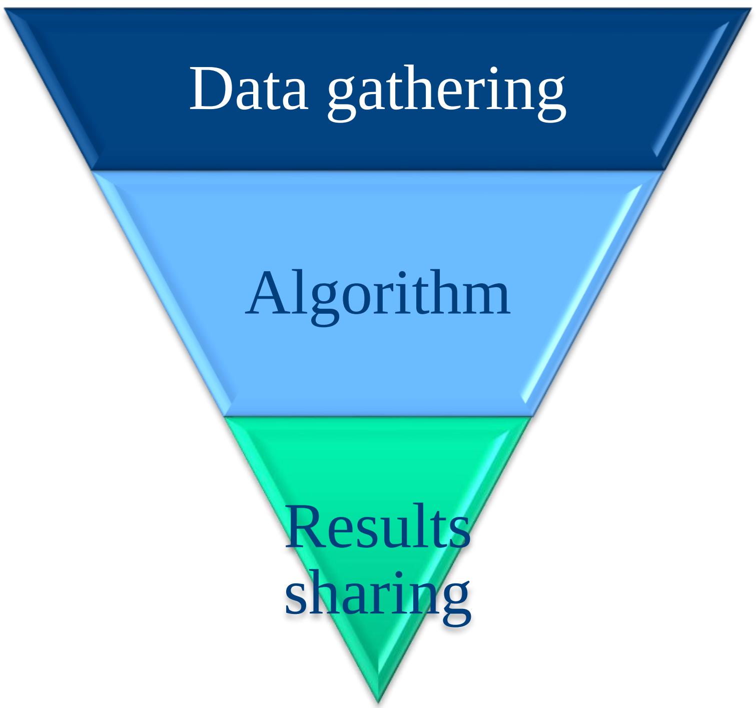
- Real example:



Source: Bloomberg, own calculation

- **Hypothesis**: Had investors followed entropy instead of std. dev, VaR or CVaR, sell-off of Russian equities would not have been so rapid after Crimea crisis outbreak





From raw Excel data…

```
T = readtable
    (inDataFileName,'sheet',sheetName);
TT = table2array(T(:,1));

TT(k,3) = std(X); %Standard Deviation
TT(k,4) = prctile(X,100 - alpha); %VaR
TT(k,5) = mean(X(X <= TT(k,4))); %CVaR
TT(k,6) = EntropyEstimationKL(X); %Entropy estim.

writetable(timetable2table(T), outDataFileName,
   'FileType','spreadsheet','Sheet','RiskMetrics')

h = plot(T.Time, table2array(T(:,3:end-1)));
savefig([outFigName, '.fig']); %MatLab figure
print([outFigName, '.png'],'-dpng'); %PNG file
```

Appendix C

…to graphic output for sharing

Conclusion

- Entropy is measure of disorder, surprise and information content
- Can therefore be used as measure of market risk
- Main advantages of entropy are:
  - Independence of underlying distribution
  - Always exists (stdDev does not)
  - Able to measure diversification correctly (VaR does not)
  - Easily to interpret as “level of surprising results”
  - More robust than e.g. standard deviation
- **Hypothesis:** using entropy can lead to calmer markets

References

1. Ormos M., Zibriczky D.: *Entropy-Based Financial Asset Pricing*, 2014, PLoS ONE 9(12)

2. Bentes S.R., Menezes R.: *Entropy: A new measure of stock market volatility?*, 2012, J. Phys.: Conf. Ser. 394

3. Pele D.T. et al.: *Information Entropy and Measures of Market Risk*, Entropy, 2017, 19(5), 226

4. Pierce J.R: Introduction to Information Theory: Symblos, Signals and Noise, 2nd Ed., 1980

5. Kozachenko L.F., Leonenko N.N.,: О статистической оценке энтропии случайного вектора, 1987, Проблемы передачи информации, 23(2)

6. Fermi E.: Thermodynamics, 1937

7. Van Ness, H.C: Understanding Thermodynamics, 1983

Martin Vesely
Risk Management Department
Czech National Bank

martin.vesely@cnb.cz
www.cnb.cz

Appendix A: Shannon entropy is subadditive risk measure (1)

- Consider two portfolios A and B with *m* and *n* possible P/L outcomes:

$$A = \begin{pmatrix} A_1 & \dots & A_m \\ p_1{}^A & \dots & p_m{}^A \end{pmatrix} \qquad B = \begin{pmatrix} B_1 & \dots & B_n \\ p_1{}^B & \dots & p_n{}^B \end{pmatrix}$$

- Combine these two portfolios into one A∗B, there are *mn* possible P/L outcomes (intersect):

$$A * B = \begin{pmatrix} A_1 + B_1 & \dots & A_1 + B_n & A_2 + B_1 & \dots & A_2 + B_n & \dots & \dots & A_m + B_n \\ \pi_{11} & \dots & \pi_{1n} & \pi_{21} & \dots & \pi_{2n} & \dots & \dots & \pi_{mn} \end{pmatrix}$$

- Entropy of combination is

$$H(A * B) = -\sum_{i=1}^{m}\sum_{j=1}^{n} \pi_{ij} \log_a \pi_{ij} = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_j{}^B p_i{}^{A|B=j} \log_a \left[p_j{}^B p_i{}^{A|B=j} \right]$$

where $p_i{}^{A|B=j}$ is conditional probability of i-th outcome in portfolio A if j-th outcome in B occurred simultaneously

Appendix A: Shannon entropy is subadditive risk measure (2)

- After some algebra it is possible to write

$$H(A*B) = H(A|B) + H(B)$$

where

$$H(A|B) = -\sum_{j=1}^{n} p_j^{\;B} \sum_{i=1}^{m} p_i^{\;A|B=j} \log_a \left[p_i^{\;A|B=j} \right]$$

Conditional entropy of A
(entropy of A if B influences A)

Partial entropy of A
(entropy of A if j-th outcome in B occurred)

- Naturally, if A and B are independent:
  1. $H(A|B) = H(A)$
  2. Any knowledge about B cannot be used for decreasing uncertainty in A, so H(A) should intuitively be maximum of H(A|B)

Note: this statement can be proven rigorously with Jensen inequality

Appendix A: Shannon entropy is subadditive risk measure (3)

- Combining…
  1. $H(A * B) = H(A|B) + H(B)$
  2. $H(A|B) \leq H(A)$
- …we have $\boldsymbol{H(A * B) \leq H(A) + H(B)}$
- **This means that combination of two portfolios together reduces risk**
  ⇒ entropy is subadditive risk measure
- This can be proven for continuous distributions as well

Appendix B: Other types of entropy

- Tsallis entropy:

$$H_T = \frac{k}{q-1}\left(1 - \sum_{i=1}^{n} {p_i}^q\right)$$

- Renyi entropy:

$$H_R = \frac{k}{1-q}\log_a\left(\sum_{i=1}^{n} {p_i}^q\right)$$

- For *q* approaching 1, both of them become Shannon entropy

- For q = 0 Renyi entropy becomes entropy in thermodynamics sense (Clausius/Boltzman definition):

$H = k\log_a n$ or $S = k\log_a \Omega$ in thermodyn. notation

Appendix C: MatLab code for Kozachenko-Leonenko estimator

```
function E = EntropyEstimationKL(input)
  n = length(input);

  %distance to the nearest neighbour
  input = sort(input);
  r = zeros(1,n);
  r(1) = input(2)-input(1);
  r(2:end-1) = min(input(3:n) - input(2:n-1), input(2:n-1) - input(1:n-2));
  r(n) = input(n)-input(n-1);

  %elimination of denegerated values
  r(r==0) = 1/sqrt(n);

  %actual estimation of entropy
  E = (1/n)*sum(log(r))+log(2*(n-1))+0.5772156649;
```

Disclaimer

- This presentation expresses the opinions of the presenter, which are not necessarily the same as those of the CNB
- This presentation is for educational purposes
- This presentation is neither an offer nor a solicitation to sell or buy any security
- This presentation is not investment advice
- The CNB and the presenter shall not be responsible for any problems, losses or other negative issues caused by the implementation and use of any method or approach described in this presentation