








Statistics of Financial Markets

Seminar 6: Model selection and risk management



Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

Acronyms used in this material

- **AD**: Anderson–Darling (test)
- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **BET**: Bucharest Exchange Trading (index)
- **BIC**: Bayesian Information Criterion
- **CDF**: Cumulative Distribution Function
- **EDF**: Empirical Distribution Function
- **EODHD**: EOD Historical Data (market-data provider)
- **ES**: Expected Shortfall
- **GED**: Generalised Error Distribution
- **KS**: Kolmogorov–Smirnov (test)
- **LR**: Likelihood Ratio
- **ML**: Maximum Likelihood
- **NIG**: Normal Inverse Gaussian (distribution)
- **QQ**: Quantile–Quantile (plot)
- **SFM**: Statistics of Financial Markets

Today's Question and Seminar Structure

- Question of the seminar
  - **among several candidate distributions for daily returns, which one should we choose, and how do we know it is good enough for VaR 1%?**
- Outputs to produce
  - the answers to the A exercises
  - the table or figure each B exercise asks for
  - a short argued answer to C1; the errors found in the AI answer of C2 and the corrected answer
  - the interpretation sentences each exercise asks for
- Nothing is handed in and nothing is graded: the seminar is for practice
- Notebook for today: open the seminar notebook in Google Colab
  - each task names its notebook section
- Exercises, their types and their order: the Exercise Map (next slide)
- Contact: Prof. Daniel Traian Pele, danpele@ase.ro

Exercise Map

Task	Question	Type	Model
A1, A2	which model has the lowest AIC and BIC, and how strong is the evidence?	Solved, Proposed	A1
A3, A4	is the larger of two nested models significantly better?	Solved, Proposed	A3
A5, A6	does a small sample pass a KS test, with known and with estimated parameters?	Solved, Proposed	A5
A7, A8	VaR 1% under two models; are the exceedances too many or too few?	Solved, Proposed	A7
B1, B2	which distribution fits BET, DAX and Bitcoin returns best?	Solved, Proposed	B1
B3, B4	do the best models pass a goodness-of-fit test? are two TLV days data errors?	Solved, Proposed	B3
B5, B6	which model gets VaR 1% right out of sample?	Solved, Proposed	B5
C1, C2	does the best model change over time? what is wrong in an AI answer?	Proposed	B1, A3

- **[Solved]**: full solution in the slides and in the notebook, a model to follow; **[Proposed]**: you solve it, following the model; the solution is discussed in class
- Part A: calculations on paper; Part B: real data in the notebook, each task ending with an interpretation question; Part C: open questions and the audit of an AI answer

Data Used

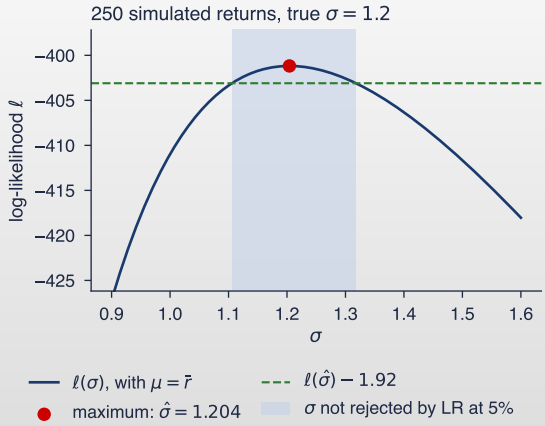
Series	Source	Price used	Period
BET, S&P 500, DAX	EODHD	close	2000–2026
Bitcoin	EODHD	close, 7 days a week	2014–2026
Banca Transilvania (TLV)	EODHD	adjusted close	2010–2026

- Daily log returns in %: $r_t = 100\,(\ln P_t - \ln P_{t-1})$; weekends and repeated holiday closes are dropped (except Bitcoin)
- In the notebook: `returns('bet')`, `fit_model('NIG', x)`; no account or key is needed

Prerequisites for Today (1/17): the Log-Likelihood

- Returns $r_1, \ldots, r_n$ i.i.d. (independent, identically distributed) with density $f(r; \theta)$
  - θ: the parameters of the model (for example μ, σ); f: the density of the model
- **Log-likelihood**: $\ell(\theta) = \sum_t \ln f(r_t; \theta)$
  - how plausible the observed returns are under the parameters θ; higher is better
  - the ML (maximum likelihood) estimate $\hat{\theta}$ maximises it; $\ell(\hat{\theta})$: the maximised value
- Normal model: $\ell(\hat{\theta}) = -\frac{n}{2}[\ln(2\pi\hat{\sigma}^2) + 1]$
  - with $\hat{\sigma}^2 = \frac{1}{n}\sum_t (r_t - \bar{r})^2$ and $\bar{r}$ the sample mean

Prerequisites for Today (2/17): the Log-Likelihood on a Chart



- Normal model with μ fixed at $\bar{r}$: ℓ as a function of σ

$$\ell(\sigma) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_t (r_t - \bar{r})^2$$

  - 250 simulated returns with true $\sigma = 1.2$
- The maximum (red point) is the ML estimate
  - $\hat{\sigma} = 1.204$: the value under which the data are most plausible
- A sharp peak means a precise estimate
  - the σ with $\ell \geq \ell(\hat{\sigma}) - 1.92$ (shaded, 1.11 to 1.32) form a 95% interval
  - $1.92 = 3.84/2$: half the 5% critical value of $\chi^2(1)$ (the LR test follows below)

SFM_ch6_seminar

Prerequisites for Today (3/17): the Candidates

- Candidates and their number of parameters k
  - Normal (2); Student-t (3: degrees of freedom ν, location, scale)
  - skewed-t (4: adds the skewness λ; $\lambda = 0$ gives the Student-t)
  - GED, generalised error distribution (3: shape β; $\beta = 2$ gives the Normal law)
  - Normal mixture of two components (5); NIG, normal inverse Gaussian (4)
- What the shape parameters do
  - a small ν or a small β: heavier tails; $\lambda < 0$: a longer left tail
- The heavier the tails of the data, the more the Normal model loses in ℓ

Prerequisites for Today (4/17): Densities of the Candidates

- **Student-t** with location m, scale s and ν degrees of freedom

$$f(r) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\,s}\left(1 + \frac{(r-m)^2}{\nu s^2}\right)^{-\frac{\nu+1}{2}}$$

  - Γ: the gamma function; the density falls like $|r|^{-(\nu+1)}$, a power tail with $\alpha = \nu$
  - variance $s^2\nu/(\nu-2)$ for $\nu > 2$; as $\nu \to \infty$, the Normal law

- **GED** with shape β

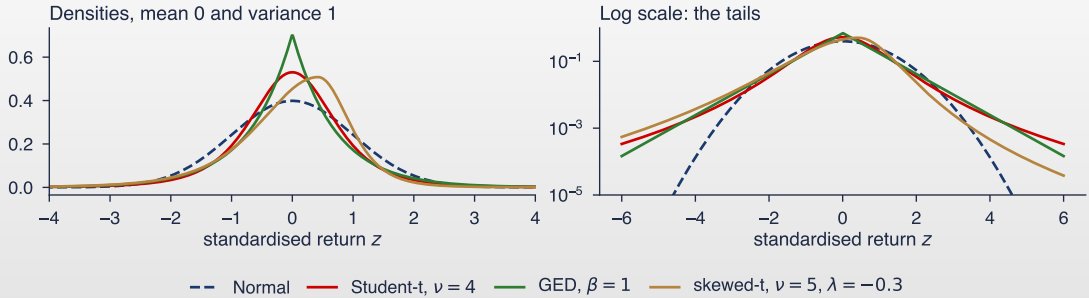
$$f(r) = \frac{\beta}{2s\,\Gamma(1/\beta)} \exp\left(-\left|\frac{r-m}{s}\right|^{\beta}\right)$$

  - $\beta = 2$: Normal; $\beta = 1$: Laplace, with a sharp peak; $\beta < 2$: tails heavier than the Normal ones

- **Normal mixture**: $f(r) = \pi\,\varphi(r;\mu_1,\sigma_1^2) + (1-\pi)\,\varphi(r;\mu_2,\sigma_2^2)$
  - $\varphi(\cdot\,;\mu,\sigma^2)$: the Normal density; five parameters $\pi, \mu_1, \sigma_1, \mu_2, \sigma_2$

- Skewed-t and NIG: an extra parameter for skewness; their densities are evaluated numerically

Prerequisites for Today (5/17): the Candidates on a Chart



- All four densities have mean 0 and variance 1: they differ in the peak and in the tails
- Left: Student-t and GED have a higher, sharper peak; the skewed-t with $\lambda < 0$ has a longer left tail
- Right (log scale): the Normal tail falls like a parabola, the others much more slowly

SFM_ch6_seminar

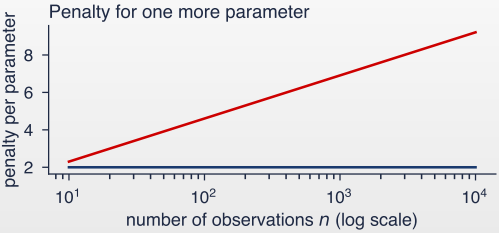
Prerequisites for Today (6/17): AIC and BIC

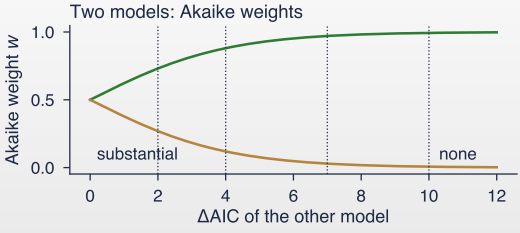
- $\mathrm{AIC} = -2\ell(\hat{\theta}) + 2k$ (Akaike information criterion)
  - fit (-2ℓ) plus a penalty of 2 for each parameter
- $\mathrm{BIC} = -2\ell(\hat{\theta}) + k \ln n$ (Bayesian information criterion)
  - penalty $\ln n$ per parameter; larger than 2 when $\ln n > 2$, i.e. $n > 7$
- Reading
  - lower is better; only differences computed on the same data matter
  - AIC is not a test: there is no p-value and no "significance"

Prerequisites for Today (7/17): ΔAIC and Akaike Weights

- $\Delta_i = \mathrm{AIC}_i - \min_j \mathrm{AIC}_j$
  - the distance of model i from the best model; $\Delta = 0$ for the winner
  - rules of thumb Burnham & Anderson (2004): $\Delta \leq 2$ substantial support; 4–7 considerably less; > 10 essentially none
- Akaike weight: $w_i = e^{-\Delta_i/2} / \sum_j e^{-\Delta_j/2}$
  - the weight of evidence for model i; the weights add up to 1
  - example: $\Delta = 0$ and $\Delta = 2$ give $w \approx 0.73$ and 0.27
- Averaged VaR: $\sum_i w_i \, \mathrm{VaR}_i$
  - the VaR of each model weighted by its Akaike weight

Prerequisites for Today (8/17): Penalties and Weights on a Chart





AIC: 2 per parameter
BIC: $\ln n$ per parameter
weight of the better model, $1/(1+e^{-\Delta/2})$
weight of the other model

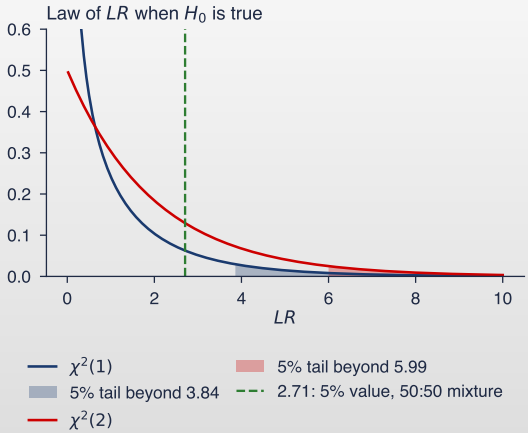
- Left: the penalty per parameter against n: 2 for AIC, $\ln n$ for BIC (6.9 for $n = 1\,000$)
- Right: two models whose AIC differ by Δ: $w_{\mathbf{better}} = 1/(1+e^{-\Delta/\mathbf{2}})$
  - dotted lines at $\Delta = 2, 4, 7, 10$: the thresholds of the rules of thumb

SFM_ch6_seminar

Prerequisites for Today (9/17): the Likelihood-Ratio Test

- Model 0 is **nested** in model 1 if it is model 1 with some parameters fixed
  - example: Student-t = skewed-t with $\lambda = 0$
- $LR = 2[\ell_{\mathbf{1}}(\hat{\theta}_{\mathbf{1}}) - \ell_{\mathbf{0}}(\hat{\theta}_{\mathbf{0}})] \approx \chi^{\mathbf{2}}(k_{\mathbf{1}} - k_{\mathbf{0}})$ under $H_{\mathbf{0}}$ (Wilks)
  - $H_{\mathbf{0}}$: the smaller model 0 is true; $k_{\mathbf{1}} - k_{\mathbf{0}}$: the number of fixed parameters
  - 5% critical values: $\chi^{\mathbf{2}}_{\mathbf{0.95}}(1) = 3.84$, $\chi^{\mathbf{2}}_{\mathbf{0.95}}(2) = 5.99$; reject $H_{\mathbf{0}}$ above them
- **Boundary**: Normal inside Student-t means $1/\nu = 0$, the edge of the parameter space
  - LR then follows a 50:50 mixture of 0 and $\chi^{\mathbf{2}}(1)$ Self & Liang (1987): 5% critical value 2.71
  - its p-value is half of the one computed with $\chi^{\mathbf{2}}(1)$
- Non-nested pairs (Student-t and GED): no LR test; use AIC or the Vuong test
  - Vuong: $V = \sqrt{n}\,\bar{m}/s_m$, $m_t = \ln f_{\mathbf{1}}(r_t) - \ln f_{\mathbf{2}}(r_t)$, with $\bar{m}$, s_m the mean and s.d. of m_t; $|V| > 1.96$: one model is closer

Prerequisites for Today (10/17): the Chi-Square Law of LR

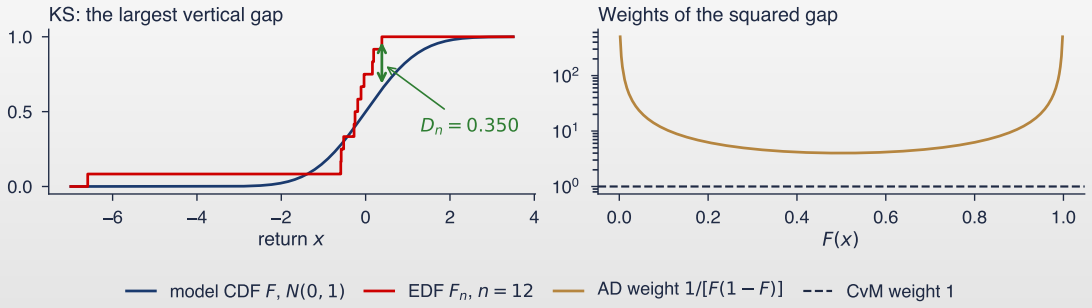


- $\chi^2(d)$, with $d = k_1 - k_0$: the law of LR under H_0
  - the law of a sum of d squared independent $N(0,1)$ variables; its mean is d
- Shaded: the 5% tails beyond the critical values of the previous slide
- p-value: the area to the right of the observed LR
- Boundary (green): half of the time $LR = 0$, so the 5% value is the 90% quantile of $\chi^2(1)$

Prerequisites for Today (11/17): Goodness-of-Fit Tests and Critical Values

- EDF (empirical distribution function): $F_n(x) = \frac{1}{n}\#\{t : r_t \le x\}$
  - the share of returns at most x; compared with the model distribution function F
- **KS** (Kolmogorov–Smirnov): D_n = the largest vertical gap between F_n and F
  - $D_n = \max_i \max\{i/n - u_{(i)},\ u_{(i)} - (i-1)/n\}$, with $u_{(i)} = F(r_{(i)})$ and $r_{(i)}$ the i-th smallest return
- **CvM** (Cramér–von Mises) and **AD** (Anderson–Darling) integrate the squared gap
  - AD weights it by $1/[F(1-F)]$, large where F is near 0 or 1: it looks at the tails
- Critical values of KS at 5%
  - F fully known: $n = 5$: 0.563; $n = 6$: 0.519; large n: $1.358/\sqrt{n}$
  - mean and variance estimated from the same data (Lilliefors): $n = 5$: 0.354; $n = 6$: 0.333
  - other models: parametric bootstrap (simulate from the fit, refit, recompute the statistic)

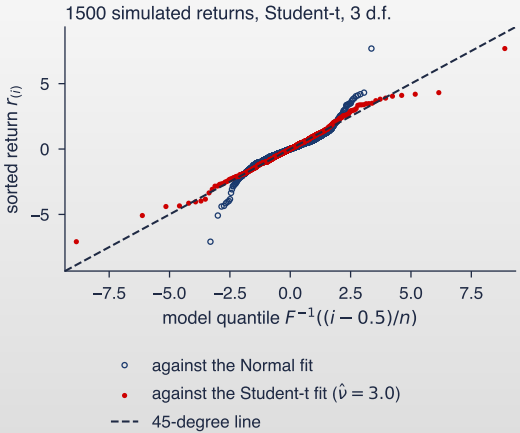
Prerequisites for Today (12/17): KS, CvM and AD on a Chart



- Left: the EDF of 12 returns against the $N(0,1)$ CDF; $D_n = \sup_x |F_n(x) - F(x)|$ is the largest vertical gap (green)
- Right: CvM, $W^2 = n \int (F_n - F)^2 \, dF$, and AD, $A^2 = n \int \frac{(F_n - F)^2}{F(1-F)} \, dF$
  - all three: large values mean a bad fit; reject when above the critical value

SFM_ch6_seminar

Prerequisites for Today (13/17): Reading a QQ Plot



- QQ plot of n returns against a fitted model F

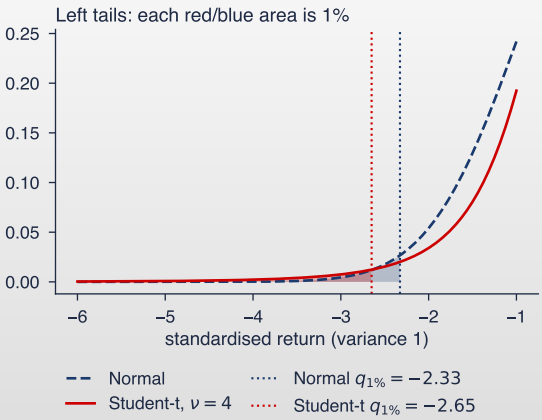
$$\left(F^{-1}\left(\frac{i-0.5}{n}\right),\ r_{(i)}\right), \quad i = 1, \ldots, n$$

  - $r_{(i)}$: the i-th smallest return
- Blue, Normal fit: an S shape; the extreme returns lie beyond the Normal quantiles: heavier tails
- Red, Student-t fit: the points stay on the line, except a few of the most extreme

Prerequisites for Today (14/17): VaR, Exceedances, Out of Sample

- $\mathrm{VaR}_{1\%} = -q_{1\%}$: the daily loss exceeded with probability 1%
  - Normal: $\mathrm{VaR}_{1\%} = -(\mu - 2.326\sigma)$; ES 2.5% (expected shortfall): the average loss beyond $q_{2.5\%}$
  - Normal ES 2.5%: $-\mu + \sigma\varphi(z_{0.025})/0.025$, with φ the $N(0,1)$ density
- **Exceedance**: a day with $r_t < -\mathrm{VaR}$; correct model: $x \sim \mathrm{Binomial}(n, 0.01)$
  - Kupiec test: $LR_{uc} = -2[\ell_0 - \ell_1]$, compared with $\chi^2(1)$
  - $\ell_0 = (n-x)\ln 0.99 + x \ln 0.01$ (rate 1%); $\ell_1 = (n-x)\ln(1-\hat{p}) + x\ln\hat{p}$ (observed rate $\hat{p} = x/n$)
- **Out of sample**: fit until a date, evaluate afterwards
  - **log score**: mean of $\ln f(r_t; \hat{\theta})$ on the test days (higher is better); it judges the whole density
  - for VaR alone: the pinball loss of the last slide of this section

Prerequisites for Today (15/17): VaR under Two Densities



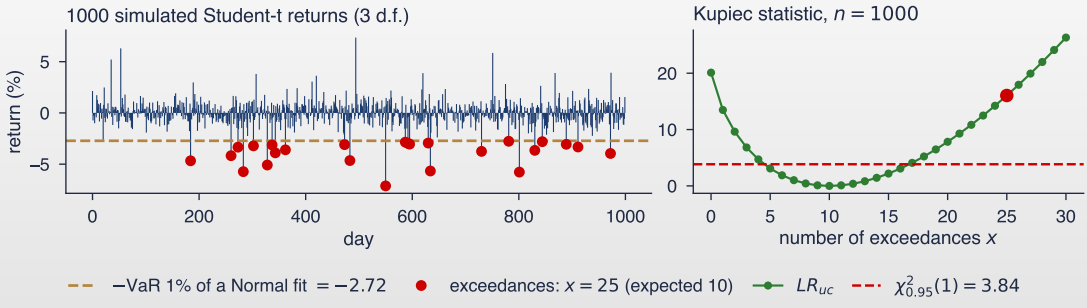
- VaR of a location-scale model

$$\mathrm{VaR}_{\mathbf{1\%}} = -(\mu + \sigma\, q^*_{\mathbf{1\%}})$$

  - σ: the standard deviation; $q^*_{\mathbf{1\%}}$: the 1% quantile of the standardised law (variance 1)
- Same variance, different tails
  - Normal: $q^*_{\mathbf{1\%}} = -2.33$; Student-t with 4 d.f. (degrees of freedom): $q^*_{\mathbf{1\%}} = -2.65$
  - for Student-t data, the Normal VaR is about 12% too low

SFM_ch6_seminar

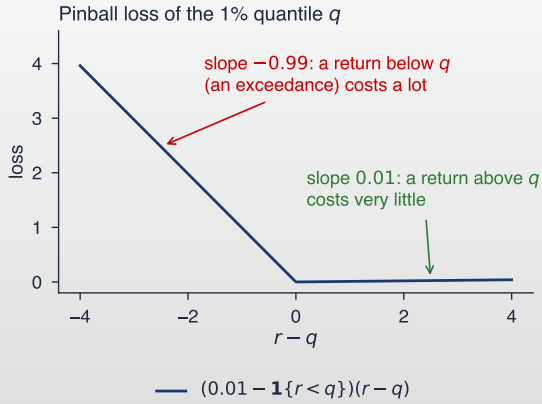
Prerequisites for Today (16/17): Exceedances and the Kupiec Test on a Chart



- Left: 1 000 Student-t returns with the VaR 1% of a Normal fit: 25 exceedances instead of the expected 10
- Right: LR_{uc} for every possible count x
  - 0 at $x = 10$; above 3.84 for $x \leq 4$ or $x \geq 17$; here $LR_{uc} = 16.0$: the Normal VaR is rejected

SFM_ch6_seminar

Prerequisites for Today (17/17): the Pinball Loss



- **Pinball loss** of a 1% quantile forecast q

$$\rho(r, q) = (0.01 - 1\{r < q\})(r - q)$$

  - $1\{\cdot\} = 1$ if the condition holds, 0 otherwise; averaged over the test days; lower is better
- A return below q costs $0.99\,|r - q|$; a return above q only $0.01\,(r - q)$
- The true 1% quantile minimises the expected loss: the pinball loss rewards a correct VaR, not a prudent one

SFM_ch6_seminar

A1: Four Models for the S&P 500 [Solved]

Task

- S&P 500, the last $n = 1000$ trading days (2022–2026); maximised log-likelihoods:
- Normal: $\ell = -1400.9$, $k = 2$
- Student-t: $\ell = -1314.7$, $k = 3$
- skewed-t: $\ell = -1313.8$, $k = 4$
- GED: $\ell = -1322.6$, $k = 3$
- 1. Compute AIC and BIC ($\ln 1000 = 6.908$) of each model.
- 2. Compute ΔAIC and the Akaike weights.
- 3. Say which model each criterion selects and how strong the evidence is.
- Report: a table with four rows and two sentences.

Solution

- 1. AIC: Normal 2805.8, Student-t 2635.4, skewed-t 2635.6, GED 2651.2; BIC: 2815.6, 2650.1, 2655.2, 2665.9
- 2. ΔAIC: 170.4, 0.0, 0.2, 15.8; $e^{-\Delta/2}$: 0.000, 1.000, 0.905, 0.000; weights 0.000, 0.525, 0.475, 0.000
- 3. AIC: Student-t and skewed-t tied ($\Delta = 0.2$); BIC: Student-t, by 5.1 points; Normal and GED have no support.

Back Exercise list

A2: Four Models for Bitcoin [Proposed]

Solution discussed in class

Task

- Bitcoin, the last $n = 1000$ days (2023–2026); maximised log-likelihoods: Model: A1.
- Normal: $\ell = -2328.0$, $k = 2$
- Student-t: $\ell = -2262.0$, $k = 3$
- NIG: $\ell = -2258.4$, $k = 4$
- Normal mixture: $\ell = -2262.3$, $k = 5$
- 1. Compute AIC, BIC, ΔAIC and the Akaike weights.
- 2. Say whether AIC and BIC select the same model.
- 3. Explain why the Normal mixture, with the most parameters, does not win.
- Report: a table with four rows and two sentences.

Back Exercise list

A3: Is Skewness Needed? Is the Normal Law Enough? [Solved]

Task

- Use the log-likelihoods of A1 (S&P 500, $n = 1000$).
- 1. Test the Student-t (H_0: $\lambda = 0$) against the skewed-t with a likelihood-ratio test at 5%.
- 2. Test the Normal law (H_0: $\beta = 2$) against the GED.
- 3. Compare the LR decision of step 1 with the AIC verdict of A1.
- Report: two LR values, two decisions and one sentence.

Solution

- 1. $LR = 2(-1313.8 - (-1314.7)) = 1.8 < 3.84$, $p = 0.18$: do not reject $\lambda = 0$
- 2. $LR = 2(-1322.6 - (-1400.9)) = 156.6$, $p < 0.001$: reject the Normal law
- 3. Consistent: AIC finds the two t-models tied; the LR test finds no significant gain from λ. For nested pairs, AIC prefers the larger model exactly when $LR > 2$ per extra parameter.

Back Exercise list

A4: A Test on the Boundary [Proposed]

Solution discussed in class

Task

- Use the log-likelihoods of A2 (Bitcoin, $n = 1000$). Model: A3.
- 1. Compute LR for the Normal law inside the Student-t.
- 2. Explain why the null law is not $\chi^2(1)$, and give the correct 5% critical value.
- 3. Say whether an LR test can compare the Student-t with the NIG.
- Report: one LR value, one critical value and two sentences.

A5: KS on Five Returns [Solved]

Task

- Daily returns (%): -2.1, -0.4, 0.3, 0.9, 1.6.
- 1. Compute D_5 against $N(0,1)$ and compare it with the 5% critical value 0.563.
- 2. Estimate $\hat{\mu}$ and $\hat{\sigma}$ (divisor n), compute D_5 against $N(\hat{\mu}, \hat{\sigma}^2)$ and compare it with the Lilliefors value 0.354.
- 3. Explain why the Lilliefors critical value is smaller.
- Report: two values of D_5, two decisions, one sentence.

Solution

- 1. $u_{(i)} = \Phi(r_{(i)})$: 0.018, 0.345, 0.618, 0.816, 0.945; the largest gap $D_5 = 0.218 < 0.563$: do not reject
- 2. $\hat{\mu} = 0.060$, $\hat{\sigma} = 1.266$; $z_{(i)}$: -1.71, -0.36, 0.19, 0.66, 1.22; $u_{(i)}$: 0.044, 0.358, 0.575, 0.747, 0.888; $D_5 = 0.175 < 0.354$: do not reject
- 3. The fitted F is pulled towards the data, so D_n is smaller under H_0; the critical value must be smaller too.

Back Exercise list

A6: KS with One Large Loss [Proposed]

Solution discussed in class

Task

- Daily returns (%): -3.0, -0.6, -0.2, 0.1, 0.5, 1.0. Model: A5.
- 1. Compute $D_{\mathbf{6}}$ against $N(0,1)$; the 5% critical value is 0.519.
- 2. Compute $D_{\mathbf{6}}$ against $N(\hat{\mu}, \hat{\sigma}^{\mathbf{2}})$; the Lilliefors value is 0.333.
- 3. The -3% day has $\Phi(-3) = 0.001$; explain why KS hardly reacts to it, while AD would.
- Report: two values of $D_{\mathbf{6}}$, two decisions, one sentence.

A7: VaR 1% under Two Models [Solved]

Task

- Daily returns (%): Normal with $\mu = 0.05$, $\sigma = 1.2$; Student-t with $\nu = 4$, location 0.06, scale 0.8; $t_{\mathbf{0.01}}(4) = -3.747$.
- 1. Compute VaR 1% under each model.
- 2. A bank uses the Normal VaR and observes 20 exceedances in 1000 days; compute the Kupiec statistic.
- 3. Decide at 5% and interpret.
- Report: two VaR values, one statistic, one decision.

Solution

- 1. Normal: $-(0.05 - 2.326 \times 1.2) = 2.742\%$; Student-t: $-(0.06 - 3.747 \times 0.8) = 2.938\%$
- 2. $\ell_{\mathbf{0}} = 980 \ln 0.99 + 20 \ln 0.01 = -101.95$; $\ell_{\mathbf{1}} = 980 \ln 0.98 + 20 \ln 0.02 = -98.04$; $LR_{uc} = -2(\ell_{\mathbf{0}} - \ell_{\mathbf{1}}) = 7.83$
- 3. $7.83 > 3.84$ ($p = 0.0051$): reject; twice the expected 10 exceedances: the Normal VaR is too low, as the thin tail predicts.

Back Exercise list

A8: Too Few Exceedances, and ES [Proposed]

Solution discussed in class

Task

- Another bank observes 4 exceedances of its VaR 1% in 1000 days. Model: A7.
- 1. Compute the Kupiec statistic and decide at 5%.
- 2. Explain why too few exceedances is also a failure.
- 3. Compute ES 2.5% of the Normal model of A7 ($z_{0.025} = -1.960$, $\varphi(1.960) = 0.0584$).
- Report: one statistic, one decision, one ES value and one sentence.

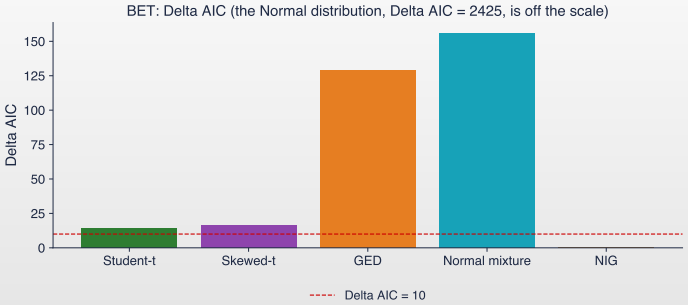
Back Exercise list

B1: Which Distribution for the BET? [Solved]

- **Question**: which of the six candidates fits the daily BET returns best, and is the skewness parameter needed?
- **Inputs**: BET closes, 2000–2026, daily log returns in % ($n = 6\,670$)
- **Tasks**
  1. Fit the six candidates by maximum likelihood and report $\ell(\hat{\theta})$ of each.
  2. Compute AIC, BIC, ΔAIC and the Akaike weights.
  3. Test the Student-t against the skewed-t with a likelihood-ratio test.
  4. Interpretation: what do the winner and the LR test say about the tails and the skewness of the BET?
- **Report**: a table with six rows, one LR test and two sentences
- Notebook: section B1

Back Exercise list

B1: Solution (1/2) [Solved]



- Winner: NIG on both criteria; Student-t $\hat{\nu} = 2.43$; skewed-t $\hat{\lambda} = -0.009$
  - $LR = 0.35$, $p = 0.555$: skewness not needed

SFM_ch6_seminar

B1: Solution (2/2) [Solved]

	$\ell(\hat{\theta})$	AIC	BIC	ΔAIC	ΔBIC	w
Normal	−11 702	23 407	23 421	2425.2	2411.6	0.000
Student-t	−10 495	20 996	21 017	14.4	7.6	0.001
skewed-t	−10 495	20 998	21 025	16.0	16.0	0.000
GED	−10 553	21 111	21 132	129.1	122.3	0.000
Normal mixture	−10 564	21 138	21 172	155.8	162.7	0.000
NIG	−10 487	20 982	21 009	0.0	0.0	0.999

- Interpretation: very heavy tails (ν close to 2) and no significant skewness; the Normal law is off by thousands of AIC points

SFM_ch6_seminar

B2: DAX and Bitcoin [Proposed]

- **Question**: do the DAX and Bitcoin choose the same distribution as the BET? Model: B1.
- **Inputs**: DAX since 2000, Bitcoin since 2014 (7 days a week), daily log returns in %
- **Tasks**
  1. Repeat steps 1–3 of B1 for each series.
  2. Say whether AIC and BIC agree for each series.
  3. Interpretation: why might the best model differ between a stock index and a crypto asset?
- **Report**: two tables and three sentences
- Notebook: section B2

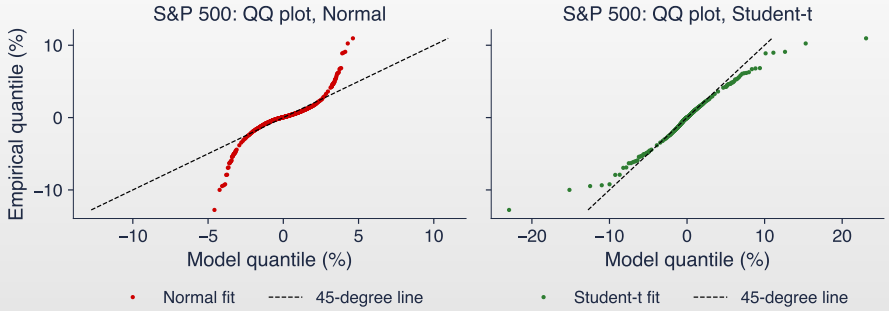
Back Exercise list

B3: Do the Best Models Pass? [Solved]

- **Question**: do the Normal and Student-t fits of the S&P 500 pass the KS, Cramér–von Mises and Anderson–Darling tests?
- **Inputs**: S&P 500 daily log returns since 2000 ($n = 6\,714$)
- **Tasks**
  1. Fit both models and compute the three statistics.
  2. Compute the naive KS p-value, which treats the parameters as known.
  3. Compute parametric bootstrap p-values with $B = 99$ (simulate, refit, recompute).
  4. Draw the QQ plot of the data against both fits.
  5. Interpretation: if both models are rejected, what is the test still good for?
- **Report**: a table of six statistics with p-values, the QQ plot and two sentences
- Notebook: section B3

Back Exercise list

B3: Solution (1/2) [Solved]



- QQ plots of the S&P 500 returns against the Normal and the Student-t fits: the Normal fails in both tails

SFM_ch6_seminar

B3: Solution (2/2) [Solved]

	KS	CvM	AD	KS naive p
Normal	0.093	22.84	127.35	< 0.001
bootstrap 95% critical value	0.011	0.12	0.75	
Student-t	0.018	0.57	4.69	0.021
bootstrap 95% critical value	0.009	0.08	0.65	

- Bootstrap $p = 0.01$ for every statistic and both models (the smallest possible with $B = 99$): both rejected
  - the naive KS p of the Student-t is far too large
- Interpretation: with $n = 6\,714$ any i.i.d. model is rejected
  - the statistics still rank the models (Student-t AD 4.69 vs 127.35)
  - the QQ plot shows where each model fails

SFM_ch6_seminar

B4: Crash or Data Error? Banca Transilvania [Proposed]

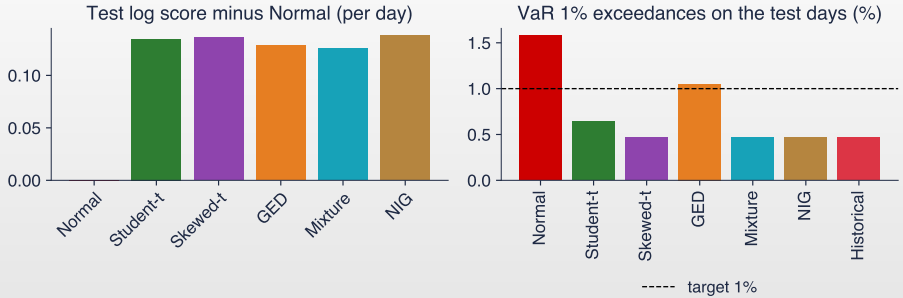
- **Question**: which of the largest TLV returns are real, and does removing the wrong ones change the chosen model? Model: B1.
- **Inputs**: TLV adjusted close since 2010, daily log returns in %; the close (not adjusted) for comparison; the BET on the same days
- **Tasks**
  1. List the three largest absolute returns with their dates.
  2. For each, compare the close and the adjusted close around the date, and the BET return of the same day.
  3. Decide which days are data errors and remove only those.
  4. Refit five candidates; compare the kurtosis, the Student-t $\hat{\nu}$, the AD statistic and the AIC winner before and after.
  5. Interpretation: what would have happened if you had removed all three days?
- **Report**: a table of three days with your verdict, a before/after table and two sentences
- Notebook: section B4

B5: VaR 1% Out of Sample for the DAX [Solved]

- **Question**: does the model chosen by AIC on 2000–2019 also give the best VaR 1% on 2020–2026?
- **Inputs**: DAX daily log returns: estimation 2000–2019 ($n = 5\,075$), test 2020–2026 ($n = 1\,711$)
- **Tasks**
  1. Fit the six candidates on the estimation days and rank them by AIC.
  2. Compute each model's mean log score on the test days.
  3. Compute each model's VaR 1%, count its exceedances on the test days and run the Kupiec test.
  4. Add historical simulation (the 1% quantile of the estimation days) and the VaR averaged with the Akaike weights.
  5. Interpretation: which model would you choose for VaR 1%, and why?
- **Report**: one table, the chart and two sentences
- Notebook: section B5

Back Exercise list

B5: Solution (1/2) [Solved]



- Left: test log score minus that of the Normal model; right: VaR 1% exceedance rates on the test days (target 1%)

SFM_ch6_seminar

B5: Solution (2/2) [Solved]

	log score	VaR 1%	exceed.	Kupiec p	pinball $\times 100$
Normal	−1.668	3.37	27	0.027	5.37
Student-t	−1.533	4.09	11	0.112	5.23
skewed-t	−1.532	4.38	8	0.013	5.36
GED	−1.539	3.91	18	0.830	5.21
Normal mixture	−1.542	4.39	8	0.013	5.37
NIG	−1.530	4.38	8	0.013	5.36
historical	–	4.33	8	0.013	5.33
averaged	–	4.37	8	0.013	5.36

- AIC on 2000–2019: NIG; best test log score: NIG; expected exceedances 17.1; lowest pinball loss: GED
- Interpretation: the AIC winner gives the best density but too few exceedances (too prudent)
  - for VaR 1% alone the GED and the Student-t were closer to 1%

SFM_ch6_seminar

B6: VaR 1% Out of Sample for Bitcoin [Proposed]

- **Question**: does the B5 conclusion hold for Bitcoin? Model: B5.
- **Inputs**: Bitcoin: estimation 2014–2019 ($n = 1\,931$), test 2020–2026 ($n = 2\,453$)
- **Tasks**
  1. Repeat steps 1–4 of B5.
  2. Compare the empirical VaR 1% of the test days with that of the estimation days.
  3. Interpretation: why do all heavy-tailed models fail the Kupiec test here, and in which direction?
- **Report**: one table and two sentences
- Notebook: section B6

C1: Does the Best Model Change over Time? [Proposed]

- **Question**: is the AIC winner of the S&P 500 and the BET the same in every two-year window?
- **Inputs**: daily log returns since 2000, non-overlapping two-year windows (2000–2001, 2002–2003, ...); candidates: Normal, Student-t, skewed-t, GED, NIG; models: B1, A1
- **Tasks**
  1. In each window, find the AIC winner, its Akaike weight and the ΔAIC of the runner-up.
  2. Count how often each model wins, for each index.
  3. Propose one explanation for the changes and one way to test it.
  4. Interpretation: should a risk model be re-selected every two years?
- **Report**: a table, a count of winners and a plan for a project
- Notebook: section C1

Back Exercise list

C2: Audit an AI Answer [Proposed]

- A student asked an AI assistant to choose a distribution for the daily S&P 500 log returns (%) since 2000 ($n = 6\,714$). The answer:
- `(a) The skewed-t has the lowest AIC, so it is significantly better than the Student-t (Delta AIC = 17.6).`
- `(b) The KS test of the Student-t fit gives p = 0.021 with scipy.stats.kstest, so the Student-t fits at the 1% level.`
- `(c) The LR test of Normal vs Student-t uses the chi2(1) critical value 3.84.`
- `(d) BIC penalises extra parameters less than AIC, because ln(n) is only 8.81.`
- `(e) The Anderson-Darling test is more sensitive than KS to the fit in the tails.`
- `(f) Since the skewed-t has the best fit, its VaR 1% is the most accurate one.`
- Tasks
  - 1. For each statement, say whether it is correct; if not, give the correct statement and, where possible, the correct number from the notebook (section C2).
  - 2. Report: a list of six verdicts with one line of justification each.

Points to Remember

- AIC and BIC rank models; $\Delta \leq 2$ is a tie; Akaike weights turn Δ into evidence
- LR tests need nested models and a parameter inside the space; on the boundary use the 50:50 mixture
- KS with estimated parameters needs Lilliefors or a bootstrap; AD sees the tails, KS does not
- Check the largest returns before modelling: remove errors, keep crashes
- The best density is not always the best VaR: validate out of sample

After the Seminar

- Lecture 6 develops each topic of today: likelihood, LR and Vuong tests, AIC and BIC, EDF tests, QQ plots, out-of-sample VaR, model uncertainty
- C1 can grow into a team project: standardised returns, more markets, rolling windows
- Reading: Burnham & Anderson (2004); Stephens (1974); Kupiec (1995)

References

- **Akaike, H. (1974).** A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- **Burnham, K.P., Anderson, D.R. (2004).** Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304.
- **Franke, J., Härdle, W.K., Hafner, C.M. (2019).** *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.
- **Hansen, B.E. (1994).** Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705–730.
- **Kupiec, P.H. (1995).** Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84.
- **Lilliefors, H.W. (1967).** On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318), 399–402.
- **Schwarz, G. (1978).** Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- **Self, S.G., Liang, K.-Y. (1987).** Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605–610.
- **Stephens, M.A. (1974).** EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347), 730–737.
- **Wilks, S.S. (1938).** The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9(1), 60–62.