CHARTA LEARNING · LESSON NOTES

Forecasting Distribution Using PIT

FRM Part II

1 A story: the model that passed

- Ashby's equity desk: a 99% VaR every evening
- The backtest counts the days the loss beat VaR
- The model passes the count
- A stressed month: big losses, and fails at 95%
- Lesson: test the whole forecast, not one point

IN THE LESSON

- Ashby's equity desk reports a ninety nine per cent V a R every evening.
- That is a loss figure the desk expects to exceed on only one day in a hundred.
- Every so often the validation team runs a backtest: it counts the days on which the actual loss was bigger than the V a R, and the model passes.
- Then a stressed month arrives.
- The losses on the bad days are far larger than the V a R, and on many ordinary bad days the desk loses more than the model's ninety five per cent figure would allow.
- The count at one confidence level looked fine.
- The rest of the forecast was wrong.
- This lesson is about a test that looks at the whole forecast.

2 What an exceedance backtest checks

- Exceedance: realised loss larger than that day's VaR
- Feature 1, frequency: about 1 − confidence, 1% at 99%
- Feature 2, independence: no clustering over time
- Kupiec tests the frequency
- Christoffersen tests independence

IN THE LESSON

- Start with the ordinary test.
- An exceedance, also called a violation, is a day on which the realised loss was larger than that day's V a R forecast.
- An accurate model shows two features.
- First, frequency.
- Exceedances should happen on a share of days close to one minus the confidence level, so about one per cent of days for a ninety nine per cent V a R.
- Second, independence.
- Exceedances should be scattered through time, with no clustering, because an exceedance today should tell you nothing about tomorrow.
- Two classic tests match the two features.
- The Kupiec test checks the frequency.
- The Christoffersen test checks independence.

3 Example 1: Ashby's frequency check

500 trading days of Ashby's 99% VaR; 6 exceedances

Expected exceedances	1% × 500	**5**
Observed exceedances	from the record	**6**
Exceedance rate	6 ÷ 500	**1.2%**
Kupiec likelihood ratio	observed versus promised rate	**0.19**
Verdict	0.19 < 3.84	**Pass**

IN THE LESSON

- Ashby's validators look at five hundred trading days of the ninety nine per cent V a R.
- The expected number of exceedances is one per cent of five hundred, which is five.
- The desk actually had six.
- The exceedance rate is six over five hundred, one point two per cent, close to one per cent.
- The Kupiec test turns this into a likelihood ratio statistic, which compares the observed rate with the promised rate.
- Here it comes to about point one nine.
- The usual five per cent critical value, from a chi squared distribution with one degree of freedom, is three point eight four.
- Point one nine is far below it, so the frequency feature passes.

4 Example 2: the clustering check

The same 6 exceedances: 4 fell on consecutive days in the stressed month

Exceedances followed by another	3 inside the run of 4	**3**
Conditional rate	3 ÷ 6	**50%**
Rate if independent	about the unconditional rate	**≈ 1%**
Verdict	50% far above 1%	**Clustered**

IN THE LESSON

- Now the second feature.
- Of Ashby's six exceedances, four came on four consecutive days in the stressed month, and the other two were isolated.
- Look at the days that followed an exceedance.
- In the run of four, three exceedance days were followed by another exceedance.
- So, of the six exceedance days, three were followed by a second one, a conditional rate of fifty per cent.
- If exceedances were independent, the chance of one tomorrow would still be about one per cent, whatever happened today.
- Fifty per cent against one per cent is clustering.
- The Christoffersen test formalises this comparison, and Ashby's model fails on independence even though its count was fine.

5 The weakness of counting exceedances

- Looks at one quantile only
- Ignores how large each exceedance was
- Uses little information: about 5 signals in 500 days
- Low power: a poor chance of catching a bad model

IN THE LESSON

- Counting exceedances has three weaknesses.
- First, it looks at one quantile only.
- A ninety nine per cent test says nothing about the ninety five per cent level, or the middle of the distribution.
- Second, it ignores how large each exceedance was.
- A loss just past the V a R and a loss three times the V a R both count as one.
- Third, it uses little information.
- With five hundred days, only about five of them carry any signal, so the test has little power, meaning a poor chance of catching a bad model.

6 Example 3: what the count misses

Desks A and B: 99% VaR $5m, 6 exceedances each; then Ashby at 95%

Desk A: average exceedance loss	just past VaR	**$5.6m**
Desk B: average exceedance loss	12 ÷ 5 = 2.4 times VaR	**$12.0m**
Ashby at 95%: expected	5% × 500	**25**
Ashby at 95%: observed	45 ÷ 500 = 9%	**45**
Kupiec at 95%	13.8 > 3.84	**Fail**

IN THE LESSON

- Here is what the count misses.
- Take two hypothetical desks, each with a ninety nine per cent V a R of five million dollars and six exceedances in five hundred days.
- On Desk A the average loss on an exceedance day was five point six million, just past the V a R.
- On Desk B it was twelve million, two point four times the V a R.
- The count rates them the same.
- Now go back to Ashby's own desk and look at its ninety five per cent V a R over the same days.
- We expect five per cent of five hundred, twenty five exceedances.
- It had forty five, a rate of nine per cent, with a Kupiec statistic of about thirteen point eight, far above three point eight four.

7 The PIT backtest in one idea

- Each day the model forecasts a whole P&L distribution
- Evening: pass the realised P&L through that day's CDF
- The result, the PIT, is a number between 0 and 1
- Correct model: PITs independent and uniform on 0 to 1

IN THE LESSON

- The probability integral transform fixes all three weaknesses at once.
- Each morning the model gives a whole forecast distribution of the day's profit and loss, not just one quantile.
- In the evening, take the realised P and L and pass it through that day's forecast distribution function.
- The answer is the probability, under the forecast, of an outcome no better than the one that happened.
- It is a number between zero and one, called the P I T of the day.
- If the model is right, the P I T values are independent and uniform on zero to one.
- So the backtest checks two things: that the values are independent, and that they are spread evenly.

8 How PITs reproduce the exceedance test

A PIT below p is an exceedance of the (1 − p) VaR

Exceedance feature	PIT version
99% frequency	Share of PITs below 0.01
95% frequency	Share of PITs below 0.05
Clustering	Low PITs following low PITs
Every level at once	Uses every observation

IN THE LESSON

- Here is the link between the two tests, row by row.
- A P I T below point zero one means the P and L was worse than the forecast's one per cent quantile, which is exactly a ninety nine per cent V a R exceedance.
- So the share of P I T values below point zero one reproduces the ninety nine per cent exceedance frequency.
- The share below point zero five reproduces the ninety five per cent frequency.
- Serial dependence among low P I T values, low values following low values, reproduces clustering.
- And the same set of P I T values answers every confidence level at once, using every one of the observations, not just the exceedances.

9 Example 4: Ashby's PITs tell both stories

Ashby: 500 PITs for the same 500 days

PITs below 0.01	6 ÷ 500	**1.2%**
PITs below 0.05	45 ÷ 500	**9.0%**
Low PITs on consecutive days	the stressed month	**4 in a row**
Same as the exceedance tests	99%, 95% and clustering	**Yes**

IN THE LESSON

- Ashby's validators compute five hundred P I T values for the same days.
- Six of them are below point zero one, a share of one point two per cent, the same as the six ninety nine per cent exceedances.
- Forty five are below point zero five, a share of nine per cent, the same as the forty five ninety five per cent exceedances.
- Four of the low values came on consecutive days, the same clustering as before.
- One set of numbers has reproduced every exceedance result.
- The validators did not need to run three separate backtests.
- And because every day gives a P I T, the values in the middle also carry information about the shape of the forecast.

10 Computing a PIT

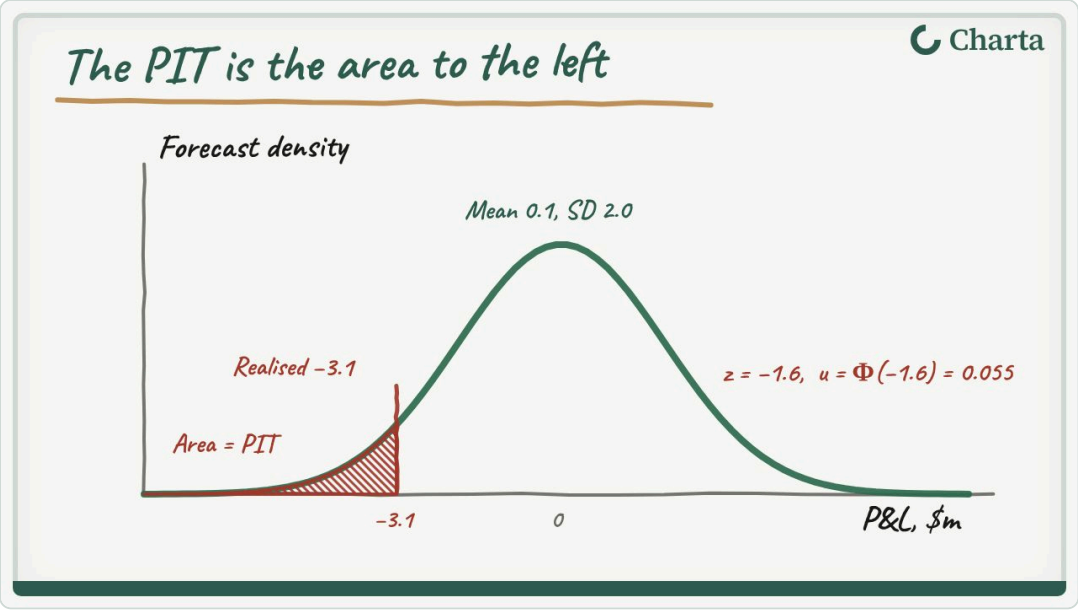
$u_t = F_t(x_t)$
Normal forecast: $z_t = (x_t - \mu_t) \div \sigma_t$, $u_t = \Phi(z_t)$

Forecast CDF F_t	Issued before the day's outcome
Realised P&L x_t	The P&L the model forecast
μ_t and σ_t	In force before the outcome
PIT u_t	A number between 0 and 1

IN THE LESSON

- Now the mechanics.
- For day t, before the outcome is known, the model issues a forecast distribution of P and L, with cumulative distribution function F t.
- Once the realised P and L, x t, is observed, the validator computes u t equals F t of x t, a number between zero and one.
- When the forecast is normal, with mean mu t and standard deviation sigma t, first standardise.
- z t is x t minus mu t, divided by sigma t.
- Then u t is phi of z t, where phi is the standard normal cumulative distribution function, the same table you use for any normal probability.
- The parameters must be the ones in force before the outcome.

11 The PIT is the area to the left



IN THE LESSON

- Here is one of Ashby's days as a picture.
- The curve is the forecast density of P and L, in millions of dollars, centred on a mean of point one with a standard deviation of two.
- The vertical red line is the realised P and L, a loss of three point one million.
- The hatched area to its left is the forecast probability of an outcome this bad or worse.
- That area is the P I T.
- Standardise first: minus three point one minus point one, divided by two, gives z of minus one point six.
- Phi of minus one point six is about point zero five five.
- So the day sat in the worst five and a half per cent of the forecast.

12 Example 5: three days of normal forecasts

Normal forecasts, $m. Day 1 and 2: mean 0.1, SD 2.0. Day 3: mean 0, SD 2.5

Day 1: loss 3.1	z = (−3.1 − 0.1) ÷ 2.0 = −1.6	**u = 0.055**
Day 2: gain 1.1	z = (1.1 − 0.1) ÷ 2.0 = +0.5	**u = 0.691**
Day 3: loss 6.0	z = (−6.0 − 0) ÷ 2.5 = −2.4	**u = 0.0082**

IN THE LESSON

- Let us compute three of Ashby's P I T values.
- Day one is the day on the chart.
- With a mean of point one and a standard deviation of two, a loss of three point one gives z of minus one point six and a P I T of point zero five five.
- Day two, same forecast, a gain of one point one million: z is plus point five, and the P I T is point six nine one.
- Day three has a new forecast: mean zero and standard deviation two point five.
- The desk loses six million.
- z is minus six over two point five, minus two point four, and the P I T is point zero zero eight two.
- Notice how each day is judged against its own forecast.

13 Example 6: a low PIT is an exceedance

Day 3: mean 0, SD 2.5, realised loss $6.0m

99% VaR on day 3	2.326 × 2.5	**$5.82m**
Day 3: loss and PIT	6.0 > 5.82; 0.0082 < 0.01	**Exceedance**
Day 1: 95% VaR	−0.1 + 1.645 × 2.0 = 3.19	**3.1 inside**
Day 1: PIT below 0.05?	0.055 > 0.05	**No**
Every level	PIT < p: a (1 − p) exceedance	**Same fact**

IN THE LESSON

- Day three shows the link to V a R.
- Its ninety nine per cent V a R was the mean minus two point three two six standard deviations, taken as a loss.
- With a mean of zero and a standard deviation of two point five, that is about five point eight two million dollars.
- The realised loss of six million is larger, so day three is a ninety nine per cent exceedance.
- Its P I T, point zero zero eight two, is below point zero one.
- The two statements are the same fact.
- Day one, by contrast, had a ninety five per cent V a R of minus point one plus one point six four five times two, about three point one nine million.
- Its loss of three point one was just inside, and its P I T of point zero five five was just above point zero five.
- The rule holds at every level: a P I T below p is an exceedance of the V a R at one minus p.

14 Example 7: PITs from historical simulation

250 scenario P&Ls from the model; count those at or below the outcome

Rule	count ≤ x_t ÷ scenarios	**PIT**
Day A: loss $4.2m	3 of 250 at or below	**u = 0.012**
Day B: gain $0.8m	180 of 250 at or below	**u = 0.72**

IN THE LESSON

- Historical simulation and Monte Carlo give a set of scenario P and L values instead of a formula.
- Then the P I T is the fraction of the scenario P and L values that are less than or equal to the realised P and L.
- Suppose Ashby's model uses two hundred and fifty scenarios.
- On one day the desk loses four point two million dollars, and only three scenarios were as bad or worse.
- The P I T is three over two hundred and fifty, which is point zero one two.
- On another day the desk gains point eight million, and one hundred and eighty scenarios were at or below that.
- The P I T is one hundred and eighty over two hundred and fifty, point seven two.

15 Rules for honest PITs

- Use the forecast in force before the outcome
- Use the P&L the model forecast: same positions, no fees
- Convention: a low PIT is a bad outcome
- Correct model: independent and uniform on 0 to 1

IN THE LESSON

- Four rules keep the P I T values honest.
- First, use the forecast in force before the outcome.
- The parameters must not be re-estimated afterwards with data that includes the day being tested.
- Second, use the P and L the model was actually forecasting.
- If the model forecasts the change in value of the positions held at the start of the day, do not compare it with a P and L that includes fees or new intraday trades.
- Third, keep the convention: a low P I T means a bad outcome, deep in the forecast's left tail.
- Fourth, remember the target.
- If the model is right, the P I T values are independent and uniform between zero and one.

16 Example 8: the look-ahead trap

Day 3 again: loss $6.0m, forecast mean 0

Forecast SD in force	z = −6.0 ÷ 2.5 = −2.4	**u = 0.0082**
SD re-estimated afterwards	z = −6.0 ÷ 3.0 = −2.0	**u = 0.023**
Exceedance at 99%?	0.023 > 0.01	**Hidden**

IN THE LESSON

- Here is what goes wrong when the first rule is broken.
- On day three Ashby's forecast standard deviation was two point five million, and the six million loss gave a P I T of point zero zero eight two, an exceedance.
- Suppose a careless validator re-estimates the volatility after the crash, using data that includes that very day, and gets three million.
- Now z is minus six over three, minus two, and the P I T becomes point zero two three.
- That is above point zero one, so the exceedance has vanished.
- The model's failure was hidden by giving it information it did not have.
- Always store the forecast distribution when it is made, and use that stored forecast.

17 Why a correct model gives uniform PITs

- Correct forecast: P(P&L below its 10% quantile) = 10%
- That is the same event as PIT below 0.1
- So P(PIT < p) = p for every p: uniform on 0 to 1
- Forecasts use all past news, so PITs are independent
- Every day lands on one common scale

IN THE LESSON

- Why should the values be uniform?
- Suppose the forecast is exactly right.
- Then the chance that the realised P and L falls below the forecast's ten per cent quantile is exactly ten per cent.
- That is the same event as the P I T falling below point one.
- The same holds for every level: the chance that the P I T is below any number p equals p.
- That is the definition of a uniform distribution on zero to one.
- And because each day's forecast already uses everything known the day before, a correct model leaves nothing to predict, so the values are also independent.
- The transform puts every day, with its own changing forecast, on one common scale.

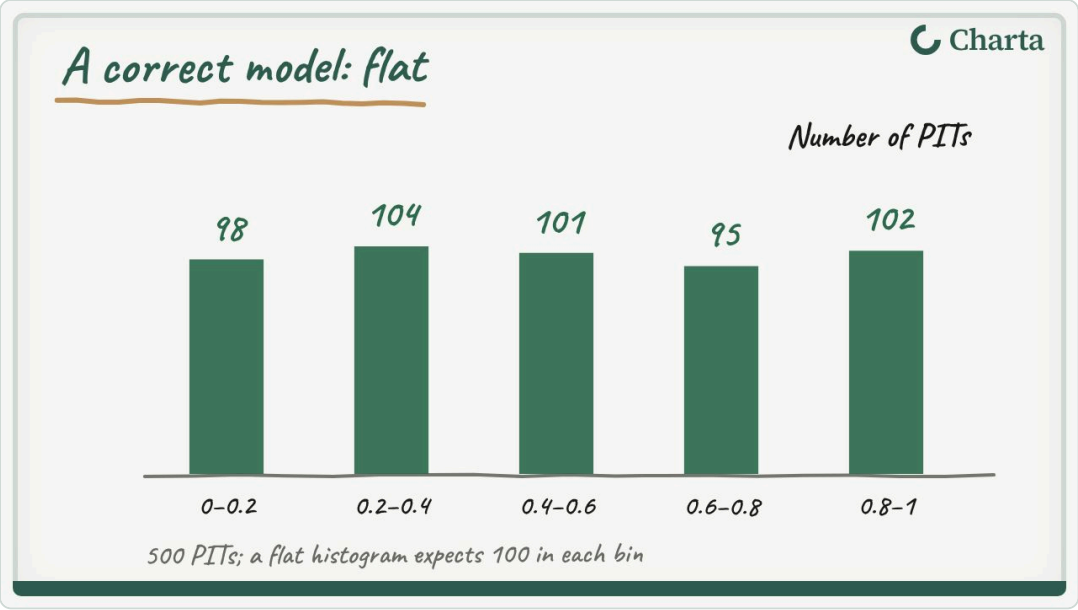
18 Reading a PIT histogram

- Five bins of width 0.2 from 0 to 1; count each bin
- Correct model: flat, about one fifth in each bin
- A failing model gives a shape that shows how it fails
- Convention: a low PIT is a bad P&L outcome
- PITs on losses: the tilts swap; U and hump do not

IN THE LESSON

- Once the values are collected, draw a histogram.
- Split zero to one into bins of equal width, here five bins of width point two, and count the values in each.
- A correct model gives a flat histogram, with every bin holding about the same share, one fifth of the values each.
- A model that fails gives a shape, and the shape says how it fails.
- We keep one convention throughout: a low P I T is a bad P and L outcome.
- If a reading defines the P I T on losses instead of P and L, a high value is the bad one, and the left and right tilts we are about to see swap over.
- The U shape and the hump read the same either way.

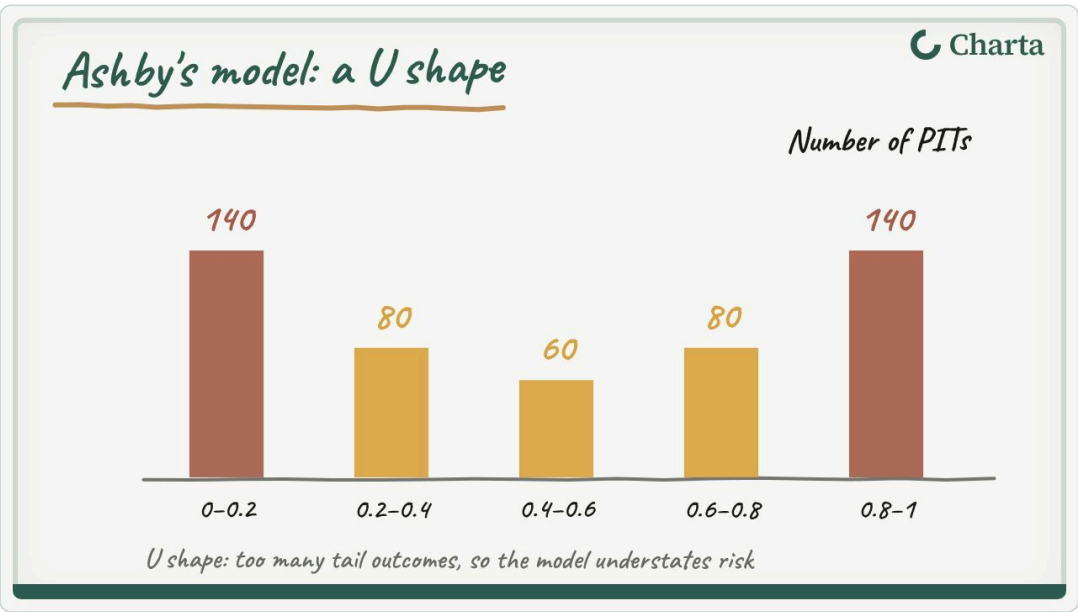
19 A correct model: flat



IN THE LESSON

- Here is the histogram from a well specified model, with five hundred P I T values.
- Under uniformity each bin should hold one fifth, one hundred values.
- The first bin, from zero to point two, holds ninety eight.
- The second holds one hundred and four.
- The middle bin holds one hundred and one.
- The fourth holds ninety five, and the last bin, from point eight to one, holds one hundred and two.
- The total is five hundred.
- The bars are not exactly equal, and they should not be, because a sample always has some noise.
- The goodness of fit tests at the end of this lesson say how much unevenness is too much.

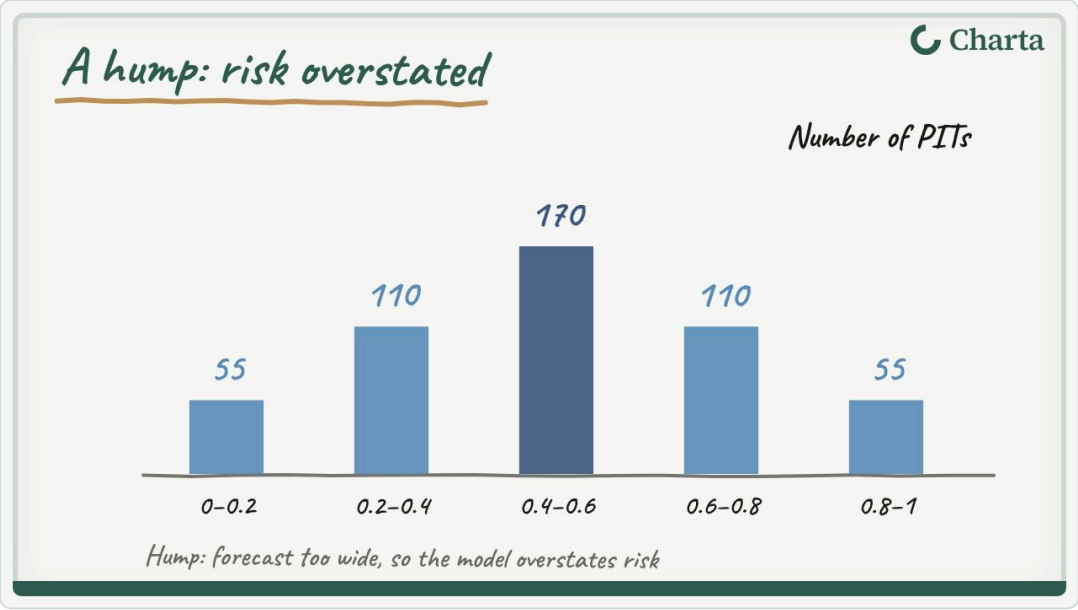
20 Ashby's model: a U shape



IN THE LESSON

- Now Ashby's own five hundred P I T values.
- The first bin holds one hundred and forty, the second eighty, the middle sixty, the fourth eighty, and the last one hundred and forty.
- The total is again five hundred, but the shape is a U, with too many values near zero and near one.
- Realised outcomes land in the tails more often than the forecast says, on both the bad side and the good side.
- The forecast distribution is too narrow.
- The model understates risk, typically because its volatility is too low or its tails are too thin.
- This matches the backtests: too many exceedances, especially at ninety five per cent.

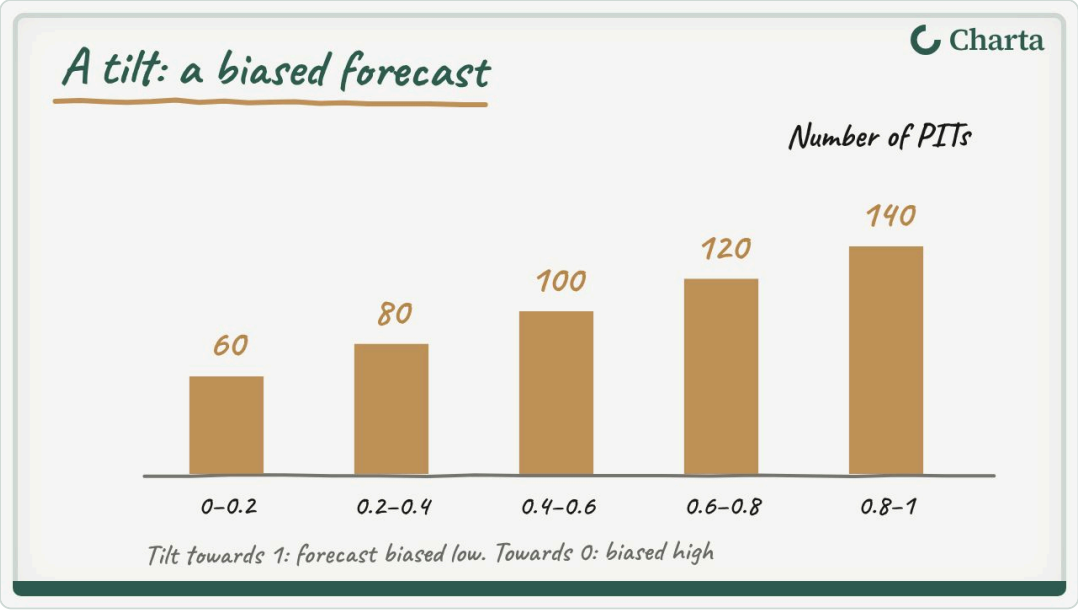
21 A hump: risk overstated



IN THE LESSON

- The opposite shape is a hump.
- Here the first bin holds fifty five, the second one hundred and ten, the middle one hundred and seventy, the fourth one hundred and ten and the last fifty five, again five hundred in all.
- There are few values in the tails and too many in the middle.
- Outcomes stay closer to the centre than the forecast expected, so the forecast distribution is too wide.
- This model overstates risk.
- It would show too few V a R exceedances, which can look reassuring, but it ties up too much capital and sets limits too tight.
- A hump is a failure too, just a conservative one.

22 A tilt: a biased forecast



IN THE LESSON

- A third failure is a tilt.
- Here the bins rise steadily: sixty, eighty, one hundred, one hundred and twenty and one hundred and forty, a total of five hundred.
- Values pile up near one, which means realised P and L keeps landing in the upper part of the forecast.
- Outcomes keep beating the forecast, so the forecast mean is biased low.
- The desk earns more on average than the model assumes.
- A tilt the other way, with values piling up near zero, means the opposite: outcomes keep falling short, and the forecast mean is biased high, too optimistic.
- With our convention, a tilt towards zero is the dangerous one for risk.

23 Example 9: flat but not independent

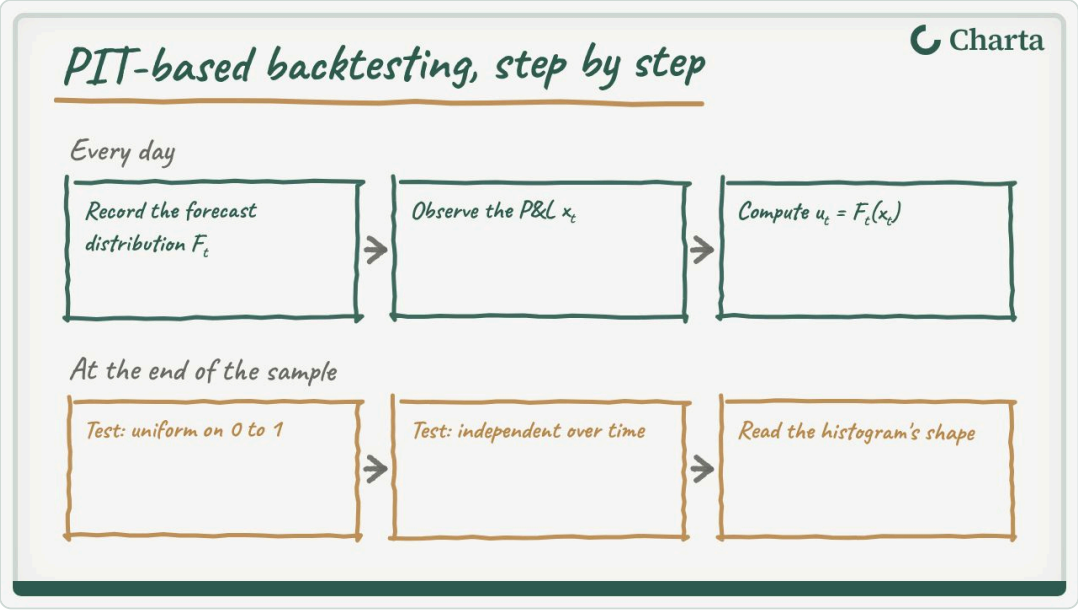
Model C: 500 PITs, a flat histogram with 100 in the lowest bin

Low PITs (below 0.2)	100 ÷ 500	**20%**
Followed by another low PIT	next day also below 0.2	**45**
Conditional rate	45 ÷ 100	**45%**
Rate if independent	same as the bin share	**20%**
Verdict	45% far above 20%	**Dependent**

IN THE LESSON

- A flat histogram is necessary but not sufficient, because the values must also be independent.
- Take a different model, Model C, with a flat histogram: one hundred of its five hundred P I T values fall in the lowest bin, below point two, exactly one fifth.
- But look at what follows them.
- Of those one hundred low days, forty five were followed the next day by another value below point two.
- That is a conditional rate of forty five per cent, against twenty per cent if the days were independent.
- Bad days bunch together, as they do when volatility rises and the model is slow to react.
- The histogram hides this completely, so always test for serial dependence as well.

24 PIT-based backtesting, step by step



IN THE LESSON

- Now we can lay out the whole procedure.
- Look at the top row, done every day.
- Record the forecast distribution before the outcome is known.
- Observe the realised P and L that the model was forecasting.
- Compute the P I T, u t equals F t of x t.
- Then the bottom row, done at the end of the sample.
- Test the null hypothesis that the values are uniform on zero to one.
- Test, separately, that they are independent over time.
- Finally, read the histogram to see how any failure arises: a U, a hump or a tilt.
- The tests say whether the model fails, and the histogram says why.

25 Comparing a distribution with the uniform

- Empirical CDF at u: share of PITs at or below u
- Uniform CDF: F(u) = u, the 45-degree line
- A correct model's empirical CDF hugs the line
- Three tests: KS, Cramér-von Mises, Anderson-Darling
- All three assume independence: test it separately

IN THE LESSON

- How do we test uniformity?
- Goodness of fit tests compare the empirical distribution function of the P I T values with the uniform one.
- The empirical distribution function at a point u is the share of the values that are at or below u.
- The uniform distribution function is simply u itself, a straight forty five degree line from zero to one.
- If the model is right, the empirical function hugs that line.
- Three tests measure the gap between them in different ways: Kolmogorov Smirnov, Cramér von Mises and Anderson Darling.
- All three assume the values are independent, which is why independence must be tested separately.

26 Three goodness-of-fit tests

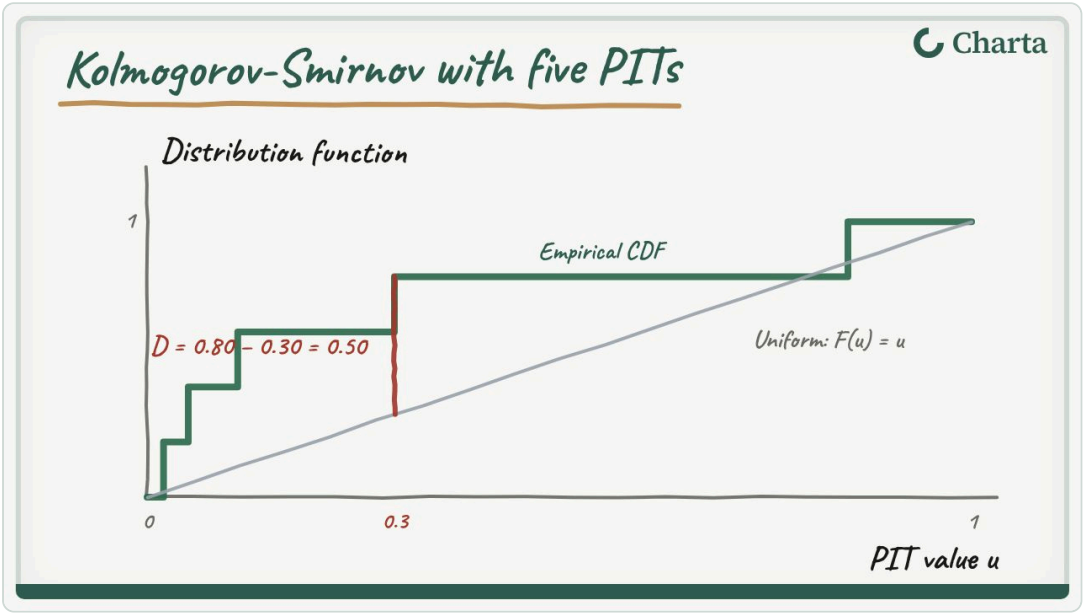
KS: D = max abs(F_n(u) − u), 5% critical ≈ 1.36 ÷ √n
CvM: n ∫ (F_n(u) − u)^2 du
AD: n ∫ (F_n(u) − u)^2 ÷ [u(1 − u)] du

Kolmogorov-Smirnov	Largest absolute gap
Cramér-von Mises	Squared gaps, equal weight
Anderson-Darling	Squared gaps, tails weighted most

IN THE LESSON

- Here are the three measures.
- Write F n of u for the empirical distribution function of n values.
- Kolmogorov Smirnov uses D, the largest absolute vertical gap between F n of u and u.
- Its approximate five per cent critical value is one point three six divided by the square root of n.
- Cramér von Mises squares the gap and sums it over the whole range from zero to one with equal weight, then multiplies by n.
- Anderson Darling also squares the gap, but weights it by one over u times one minus u.
- That weight is large near zero and near one, so Anderson Darling is the most sensitive of the three to the tails.

27 Kolmogorov-Smirnov with five PITs



IN THE LESSON

- To see the Kolmogorov Smirnov gap, take a tiny sample of five P I T values: point zero two, point zero five, point one one, point three and point eight five.
- The grey diagonal is the uniform distribution function.
- The green staircase is the empirical one: it jumps by one fifth at each value.
- Because four of the five values are low, the staircase climbs far above the line on the left.
- The largest gap is at point three.
- Just after that value the staircase stands at point eight, while the line is at point three, so the gap shown in red is point five.
- That is the statistic D.

28 Example 10: five PITs are too few

Five PITs: 0.02, 0.05, 0.11, 0.30, 0.85. Check both sides of each jump

After 0.02 and 0.05	0.20 − 0.02; 0.40 − 0.05	**0.18; 0.35**
After 0.11 and 0.85	0.60 − 0.11; 1.00 − 0.85	**0.49; 0.15**
After 0.30: the largest	0.80 − 0.30	**D = 0.50**
Before the jumps: largest	0.85 − 0.80	**0.05**
Critical value, n = 5, 5%	0.50 < exact 0.563	**Cannot reject**

IN THE LESSON

- Here is the calculation.
- Sort the values.
- At each value, check the gap just after the jump, i over five minus u, and just before it, u minus i minus one over five.
- After the jumps: at point zero two it is point two minus point zero two, point one eight.
- At point zero five, point four minus point zero five, point three five.
- At point one one, point six minus point one one, point four nine, and at point eight five, one minus point eight five, point one five.
- At point three, point eight minus point three, point five, the largest gap.
- Before the jumps the largest gap is only point eight five minus point eight, point zero five, so the gaps after the jumps are the ones that matter.
- The exact five per cent critical value for five observations is point five six three.
- D of point five is below it, so we cannot reject uniformity, even though four of five values are low.
- Five observations give almost no power.

29 Example 11: Kolmogorov-Smirnov for Ashby

Ashby: 500 PITs, bins 140, 80, 60, 80, 140; check the gap at each bin edge

Gap at 0.2	140 ÷ 500 = 0.28 vs 0.20	**0.08**
Gaps at 0.4 and 0.6	0.44 vs 0.40; 0.56 vs 0.60	**0.04 each**
Gap at 0.8	360 ÷ 500 = 0.72 vs 0.80	**0.08**
Critical value, 5%	1.36 ÷ √500	**0.061**
Verdict	0.08 > 0.061	**Reject**

IN THE LESSON

- Now Ashby's five hundred values from the U shaped histogram, checking the gap at each bin edge.
- The empirical distribution function at point two is one hundred and forty over five hundred, point two eight, against point two on the line, a gap of point zero eight.
- At point four it is two hundred and twenty over five hundred, point four four, a gap of point zero four.
- At point six it is two hundred and eighty over five hundred, point five six, a gap of point zero four the other way.
- At point eight it is three hundred and sixty over five hundred, point seven two, a gap of point zero eight, also below the line.
- So D is point zero eight.
- The critical value is one point three six over the square root of five hundred, about point zero six one.
- Point zero eight is above it, so uniformity is rejected.

30 Example 12: why Anderson-Darling watches the tails

Anderson-Darling weight: 1 ÷ [u(1 − u)]

At u = 0.5	1 ÷ (0.5 × 0.5)	**4.0**
At u = 0.05	1 ÷ (0.05 × 0.95)	**21.1**
At u = 0.01	1 ÷ (0.01 × 0.99)	**101.0**
Tail versus middle	101.0 ÷ 4.0	**≈ 25 times**

IN THE LESSON

- Kolmogorov Smirnov and Cramér von Mises treat a gap the same wherever it occurs, yet tail gaps are naturally small, because both functions are pinned near zero and one at the ends.
- Anderson Darling corrects that with its weight, one over u times one minus u.
- At u of point five the weight is one over point two five, which is four.
- At point zero five it is one over point zero four seven five, about twenty one.
- At point zero one it is one over point zero zero nine nine, about one hundred and one, roughly twenty five times the weight in the middle.
- For V a R, where the tails are what matter, Anderson Darling is the test most likely to catch a tail failure.

31 Choosing between the three tests

All three compare the PITs' empirical CDF with the uniform CDF

Test	Measures	Strength
Kolmogorov-Smirnov	Largest gap	Simple; weak in tails
Cramér-von Mises	Squared gaps, equal weight	Whole range; weak in tails
Anderson-Darling	Squared gaps, tail weight	Most tail-sensitive
All three	Assume independence	Test it separately

IN THE LESSON

- Here is how the three tests compare, row by row.
- Kolmogorov Smirnov measures the single largest gap, so it is simple, with an easy critical value, but it is least sensitive in the tails.
- Cramér von Mises sums the squared gaps with equal weight, so it uses the whole distribution, but it is also weak in the tails.
- Anderson Darling sums the squared gaps with tail weighting, so it is the most sensitive in the tails, which is where a V a R model matters most.
- The last row applies to every test.
- All three assume the values are independent, so independence must be checked separately.

32 Mistakes to avoid

- Passing at 99% says nothing about other levels or sizes
- Use the forecast in force and the P&L it forecast
- Check the convention: on losses, the tilts reverse
- Flat is not enough: test independence too
- Tests assume independence; small samples, low power

IN THE LESSON

- Let us collect the traps.
- First, a model that passes the ninety nine per cent count can still fail at other levels or in the size of its losses, so do not stop at one quantile.
- Second, never re-estimate the parameters after the fact, and always use the P and L the model forecast.
- Third, check the convention.
- With P and L, a low P I T is bad.
- With losses the tilts reverse.
- Fourth, a flat histogram is not a pass on its own, because clustered values can still be flat.
- Last, the goodness of fit tests assume independence, and with few observations they have little power, so a failure to reject is weak evidence.

33 Recap

- Kupiec: frequency. Christoffersen: clusters. Sizes ignored
- PIT: u_t = F_t(x_t); normal Φ(z_t); simulation share ≤ x_t
- Correct model: independent, uniform PITs
- U: risk understated. Hump: overstated. Tilt: biased
- KS largest gap; CvM squared gaps; AD weights the tails

IN THE LESSON

- To recap.
- An accurate V a R model has exceedances at the promised frequency and no clustering, which Kupiec and Christoffersen test.
- But counting ignores sizes.
- The P I T passes each day's realised P and L through that day's forecast distribution function.
- For a normal forecast it is phi of z, and for simulation it is the share of scenarios at or below the outcome.
- A correct model gives independent, uniform values.
- A U shape means risk is understated, a hump means it is overstated, and a tilt means bias.
- Kolmogorov Smirnov, Cramér von Mises and Anderson Darling test uniformity, with Anderson Darling watching the tails.

34 Try three quick questions

1. Normal forecast: mean $0.2m, SD $1.5m. Realised loss $2.8m. PIT?
   Answer: z = (−2.8 − 0.2) ÷ 1.5 = −2.0; Φ(−2.0) = 0.023. A 95% exceedance, not 99%.
2. A PIT histogram has a hump in the middle and few tail values. Meaning?
   Answer: Forecast too wide: the model overstates risk.
3. 400 PITs, KS statistic D = 0.08. Reject uniformity at 5%?
   Answer: Critical 1.36 ÷ √400 = 0.068. 0.08 > 0.068: reject.

IN THE LESSON

- Now try three quick questions.
- Pause the video, think, then listen.
- Question one.
- A normal forecast has a mean of point two million dollars and a standard deviation of one point five million.
- The realised P and L is a loss of two point eight million.
- What is the P I T?
- z is minus two point eight minus point two, divided by one point five, which is minus two.
- Phi of minus two is about point zero two three, so it is a ninety five per cent exceedance but not a ninety nine per cent one.
- Question two.
- A P I T histogram has a hump in the middle and few values in the tails.
- What does it mean?
- The forecast is too wide, so the model overstates risk.
- Question three.
- Four hundred P I T values give a Kolmogorov Smirnov statistic of point zero eight.
- Reject at five per cent?
- The critical value is one point three six over twenty, point zero six eight, so yes, reject.