Question 1 of 6

P2.T5.25.3.1

Based on the data provided, which portfolio's back test indicates a more accurate VaR model?

Portfolio	Exceedances Observed (% of Days)	Expected Exceedances (% of Days)	Time Between Exceedances (Days)	PIT Distribution Skewness	PIT Kurtosis
Portfolio X	1.00%	1.00%	5-12	-0.05	2.8
Portfolio Y	1.20%	1.00%	1-3, 15-20	-0.45	5.6

- **A.** Both Portfolio X and Y
- **B.** Only Portfolio Y
- **C.** Only Portfolio X
- **D.** Neither Portfolio X nor Y

Explanation

C is CORRECT.

25.3.1. C. Only Portfolio X

Unconditional Coverage Test

- Portfolio X's observed exceedances (1.00%) align exactly with expected exceedances (1.00%), meeting the unconditional coverage criterion.
- Portfolio Y's observed exceedances (1.20%) are higher than expected (1.00%), failing unconditional coverage.

Independence Test

- Portfolio X's time between exceedances (5-12 days) shows consistent intervals, suggesting no clustering.
- Portfolio Y's time between exceedances varies significantly (1-3 days followed by 15-20 days), indicating clustering during stress periods, which fails the independence criterion.

PIT Uniformity Test

- Portfolio X’s PIT skewness (-0.05) is close to 0, and kurtosis (2.8) is close to 3, indicating a near-uniform distribution and well-calibrated tails.
- Portfolio Y’s PIT skewness (-0.45) and high kurtosis (5.6) show left-skewed and heavy-tailed behaviour, indicating poor calibration.

Conclusion

- Portfolio X passes all criteria, making it the more accurate VaR model.

Question 2 of 6

P2.T5.25.3.2

Using the table below, determine the Statistic (Z-score) that corresponds to the PIT value on Day 4. Additionally, identify a day where the realized result was in the lower tail of the forecasted distribution, meaning the model expected a significantly higher value.

Expected loss	0.5
Standard Deviation	1
Acctual average of loss	0.3

Returns and PIT data		Returns and PIT data	
Realised results	**Day**	**Statistic**	**PIT value**
1.2	1	0.7	75.8%
-0.3	2	-0.8	21.2%
-0.8	3	-1.3	9.7%
2	4		93.3%
-0.7	5	-1.2	11.5%
0	6	-0.5	30.9%
1.9	7	1.4	91.9%
-1.8	8	-2.3	1.1%
0.4	9	-0.1	46.0%
1.1	10	0.6	72.6%

- **A.** 1.50, Day 4
- **B.** -5.00, Day 4
- **C.** 1.50, Day 8
- **D.** 3.40, Day 3

Explanation

C is CORRECT.

25.3.2. C. 1.50, Day 8

Part 1: Identifying the Statistic (Z-score) for Day 4

$$z = (\text{Realized Result - Expected Loss})/(\text{Standard Deviation})$$

$$z = (2 - 0.5)/1 = 1.5$$

Part 2: Identifying the Day with the Lower Tail Observation

A realized result is in the lower tail of the PIT means the observed value was lower than what the model expected. The lowest PIT values in the table are:

- Day 8: PIT = 1.1%
- Day 3: PIT = 9.7%
- Day 5: PIT = 11.5%

Day 8 has the lowest PIT value and is also the only PIT value included with the correct Z-score.

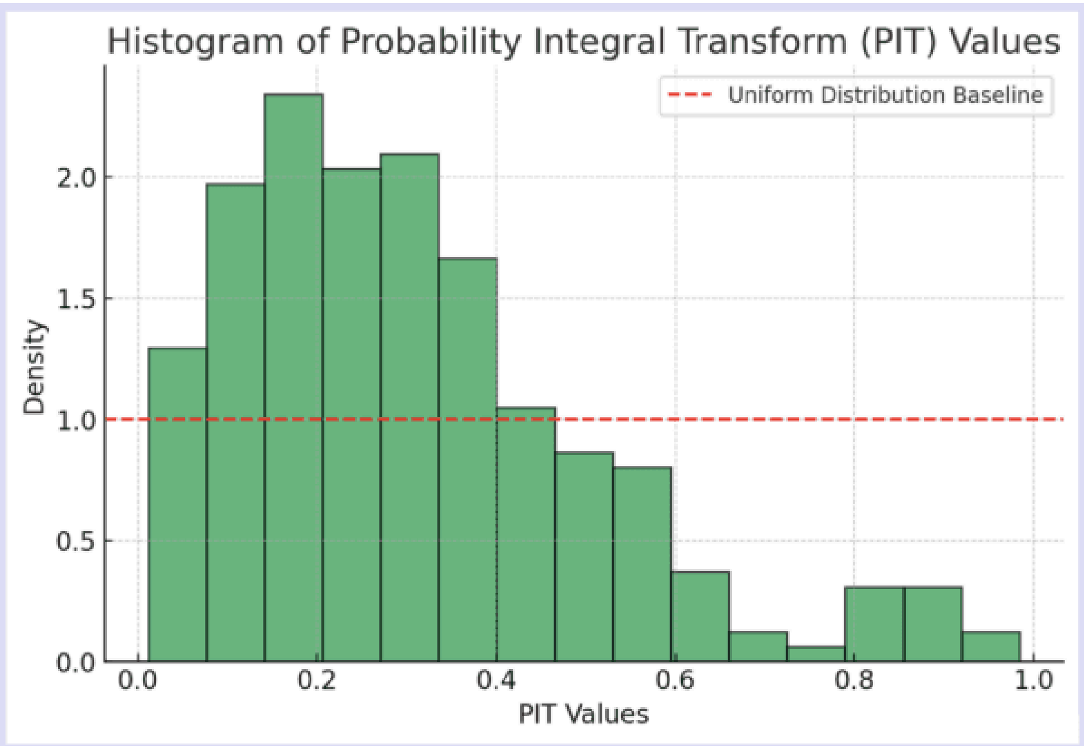
Question 3 of 6

P2.T5.25.4.1

Adam Knight, a risk manager assigned to Monument Growth Fund, is evaluating a VaR model's performance during back testing. He notices that the model reports a 99% VaR of $5 million, meaning daily losses should exceed this threshold about 1% of the time (roughly 2-3 times in a 250-day period).

However, after analyzing historical data from 2022, a period marked by aggressive central bank rate hikes and extreme market swings, Adam observes the following results:

- Over 250 trading days, actual portfolio losses exceeded the 99% VaR level 8 times instead of the expected 2-3 times.
- Most of these exceedances occurred during periods of heightened volatility in equities, fixed income, and currency markets following unexpected inflation data and Federal Reserve policy announcements.
- A review of Probability Integral Transforms (PITs) shows a concentration of values near zero, indicating a systematic underestimation of extreme losses (as illustrated in the histogram below).



Which goodness of fit test would be most appropriate to identify the issue Adam is having?

- **A.** Kolmogorov-Smirnov (KS) Test, which measures the overall deviation between the empirical and theoretical CDF, ensuring calibration across all quantiles.
- **B.** Anderson-Darling (A–D) Test, which places more emphasis on the tails of the distribution, making it particularly useful for detecting underestimation of extreme losses.
- **C.** Cramér-von Mises (CvM) Test, which balances sensitivity across the entire PIT distribution, focusing equally on the center and tails.
- **D.** Kupiec Test, which checks the frequency of exceedances but does not evaluate the distribution of PITs.

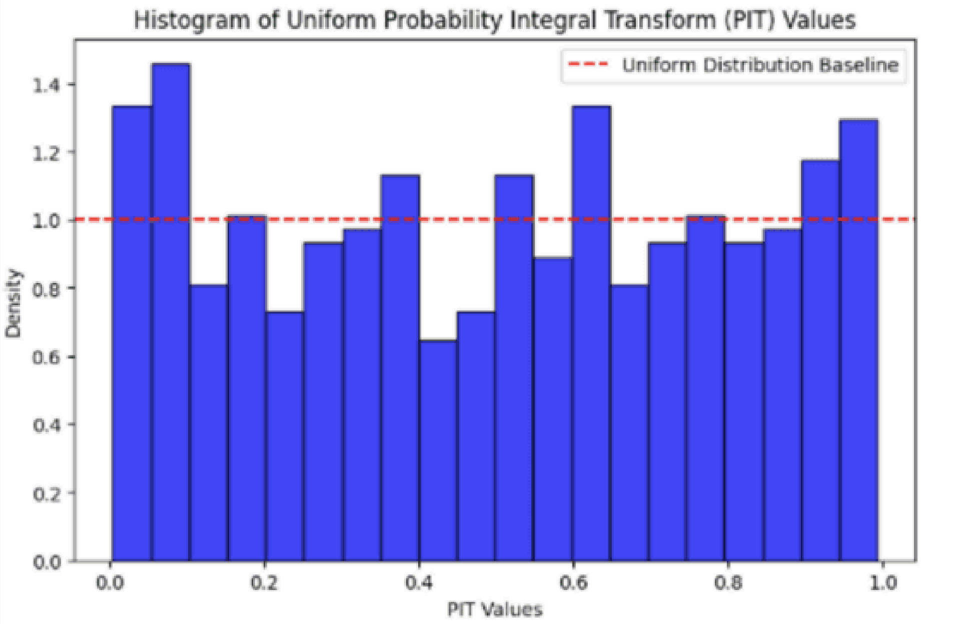
Explanation

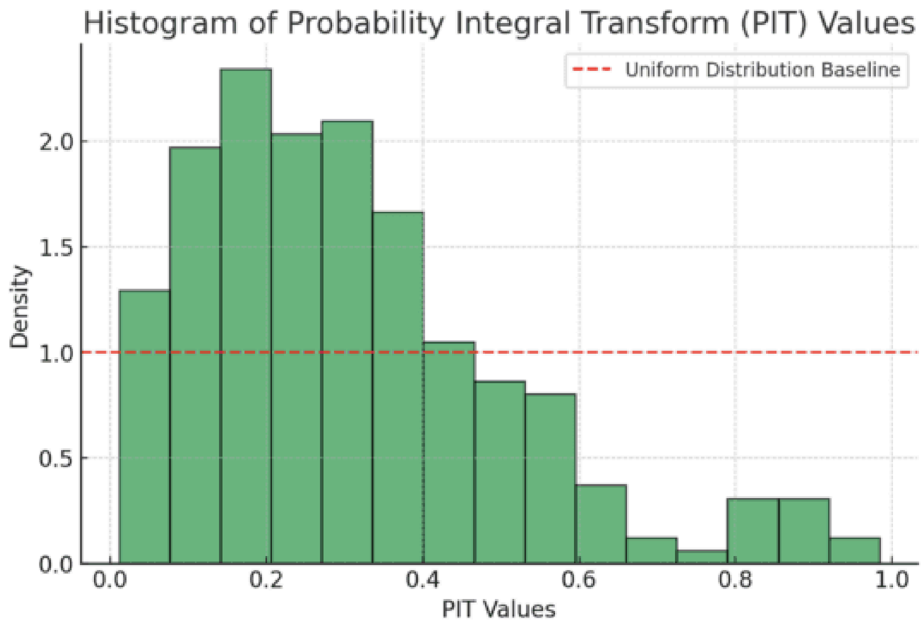
B is CORRECT.

25.4.1. B. Anderson-Darling (A–D) Test highlights tail mispricing because the test statistic exceeds the critical value, which is consistent with the clustering of PIT values near 0 and 1.

Exceedance Frequency Shows Tail Underestimation because the 99% VaR model expects 2-3 exceedances over 250 days, whereas the backtest results show 8 exceedances, indicating that the model consistently underestimates large losses—a clear tail-risk issue.

The PIT Histogram further confirms the issue as the PIT histogram in the scenario shows a concentration of values near zero rather than a uniform distribution. This indicates that extreme losses occur more frequently than predicted, confirming a problem with tail modeling.





The A–D Test is the Best Fit because it is designed for tail-risk detection. A–D test applies higher weights to deviations in the extremes, making it the best tool to detect underestimated losses in volatile conditions. Secondly, it is more sensitive than KS or CvM.

The Kolmogorov-Smirnov (KS) test and Cramér-von Mises (CvM) test assess overall fit; neither prioritizes the tails as strongly as the A–D test, making them less effective for identifying extreme loss misestimation.

Below is a summary for each test mentioned in the scenario:

Test	Purpose	Method	Advantages	Limitations	Application
Kolmogorov-Smirnov (KS)	Assess the overall fit of the PIT distribution compared to uniformity.	Measures the maximum difference between empirical and theoretical CDFs.	provides a clear measure of overall deviation.	Less sensitive to deviations in the tails.	Detect general calibration issues in the PIT distribution.
Anderson-Darling (A–D)	Focus on tail mispricing in the PIT distribution.	Emphasizes deviations in the tails through weighted squared differences.	Highly sensitive to extreme deviations; ideal for tail-risk analysis.	Lower pass rates due to strict tail emphasis.	Evaluate tail risks where accurate modelling of extremes is crucial.
Cramér-von Mises (CvM)	Provide a balanced evaluation of deviations across the entire PIT distribution.	Considers the mean squared difference across the full distribution.	Balances sensitivity to both center and tails; holistic approach.	Less specialized for tail-specific deviations.	Assess overall model calibration with equal weight to all deviations.

Question 4 of 6

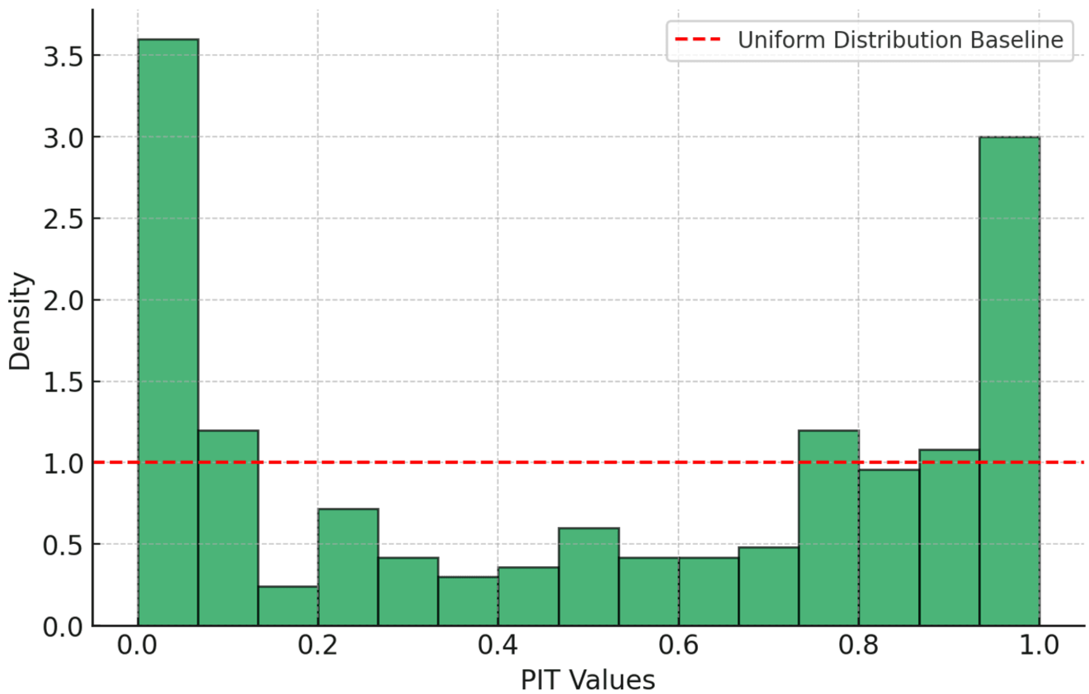
P2.T5.25.4.2

A risk Analyst is backtesting a VaR model using a 99% confidence level to evaluate PIT distributions for a portfolio over 100 days.

The test results are as follows:

	Critical Value (99%)	Observed Statistic
Kolmogorov-Smirnov	0.136	0.09
Anderson-Darling	4	4.8
Cramér-von Mises	0.115	0.075

The analyst also reviews the PIT histogram below:



Based on the test results and the histogram, which test most clearly indicates goodness-of-fit issues with the model, and why?

- [ ] **A.** Kolmogorov-Smirnov (KS) Test confirms no significant calibration issues because the test statistic is below the critical value, meaning there is no major deviation in overall fit.
- [x] **B.** Anderson-Darling (A–D) Test highlights tail mispricing because the test statistic exceeds the critical value.

- **C.** Cramér-von Mises (CvM) Test confirms the model is well-calibrated because its test statistic is below the critical value.

- **D.** All tests indicate the model is well-calibrated.

Explanation

B is CORRECT.

25.4.2. B. Anderson-Darling (A–D) Test highlights tail mispricing because the test statistic exceeds the critical value.

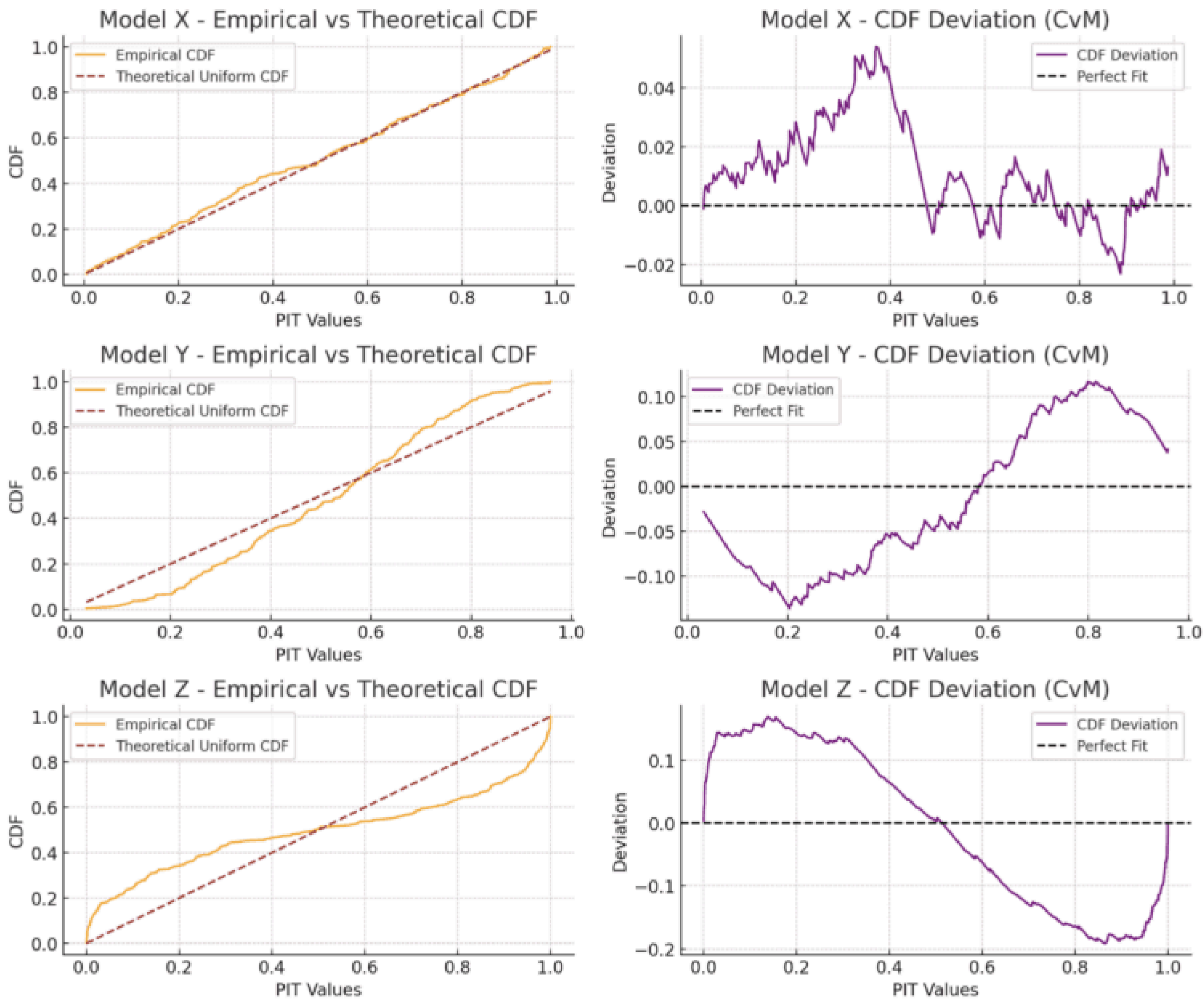
The Anderson-Darling test result (4.8>4.0) highlights a significant tail mispricing issue, making it the most relevant test in this back-testing scenario.

Test	Judgment	Interpretation
Kolmogorov-Smirnov (KS)	0.09<0.136: **Pass**	The model does not show a significant overall deviation from the expected uniform PIT distribution. However, since the KS test evaluates deviations across the entire distribution equally, it may fail to detect tail-specific issues, which are critical for risk management.
Anderson-Darling (A–D)	4.8>4.0 **Fail**	The test strongly indicates tail miscalibration, which aligns with the clustering of PITs near 0 and 1. This means the VaR model is underestimating extreme losses (left tail) and overestimating extreme gains (right tail), leading to a misleading risk assessment during market stress. In practice, this could result in inadequate capital buffers or excessive margin calls during volatile periods.
Cramér-von Mises (CvM)	0.075<0.115: **Pass**	The model does not show significant overall deviations, confirming general calibration is reasonable. However, the CvM test treats all parts of the PIT distribution equally, meaning it does not prioritize tail risks. While this suggests that the model is not entirely miscalibrated, it does not contradict the A–D test's failure in the tails, which is the more critical issue for risk managers.

Question 5 of 6

P2.T5.25.4.3

The charts below represent the Probability Integral Transform (PIT) results for three different VaR models tested using Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) goodness-of-fit tests. Based on the empirical vs. theoretical CDFs and deviation plots:



PIT Results	KS Stat	KS p-value	CvM Stat	CvM p-value
Model X	0.0538	0.4478	0.0910	0.6312
Model Y	0.1398	0.0001	1.5855	0.0001
Model Z	0.1954	0.0000	4.0694	0.0000

Which of the following statements correctly describes the observations and test outcomes?

- [x] **A.** Model X PITs are the most uniform as it has low KS and CvM statistics.
- [ ] **B.** Model Y PITs are slightly biased as it has a high CvM p-value but shows significant tail deviations, suggesting poor model calibration for extreme events.
- [ ] **C.** Model Z PITs are heavily clustered because it shows extreme deviations in the empirical CDF and high p-values for both KS and CvM tests, reflecting an overconservative risk model.
- [ ] **D.** The Model X exhibits the most tail clustering.

Explanation

A is CORRECT.

25.4.3. A. Model X PITs are the most uniform as it has low KS and CvM statistics.

Both KS and CvM tests confirm that the Perfect Uniform model aligns well with the uniform distribution. The charts show minimal deviations in both CDF and deviation plots.

- **Regarding B,** while Model Y (Slightly Biased) has moderate deviations, its CvM p-value (0.0001) is very low, not high, meaning the test rejects uniformity rather than supporting it. The description of significant tail deviations is misleading, as the test statistics reflect a broader deviation across the distribution, not just the tails.
- **Regarding C,** Model Z (Heavily Clustered) has extremely low p-values (0.0000) and high test statistics, indicating strong rejection of uniformity due to severe miscalibration. While extreme deviations are visible, the claim that it reflects an overconservative risk model is incorrect. Instead, clustering near 0 and 1 suggests underestimation of tail risks in certain areas and overestimation in others.
- **Regarding D,** Model X (which is the most uniform) does not exhibit tail clustering. If it did, we would expect higher KS and CvM statistics and lower p-values, but the flat deviation plot confirms that no significant clustering is present.

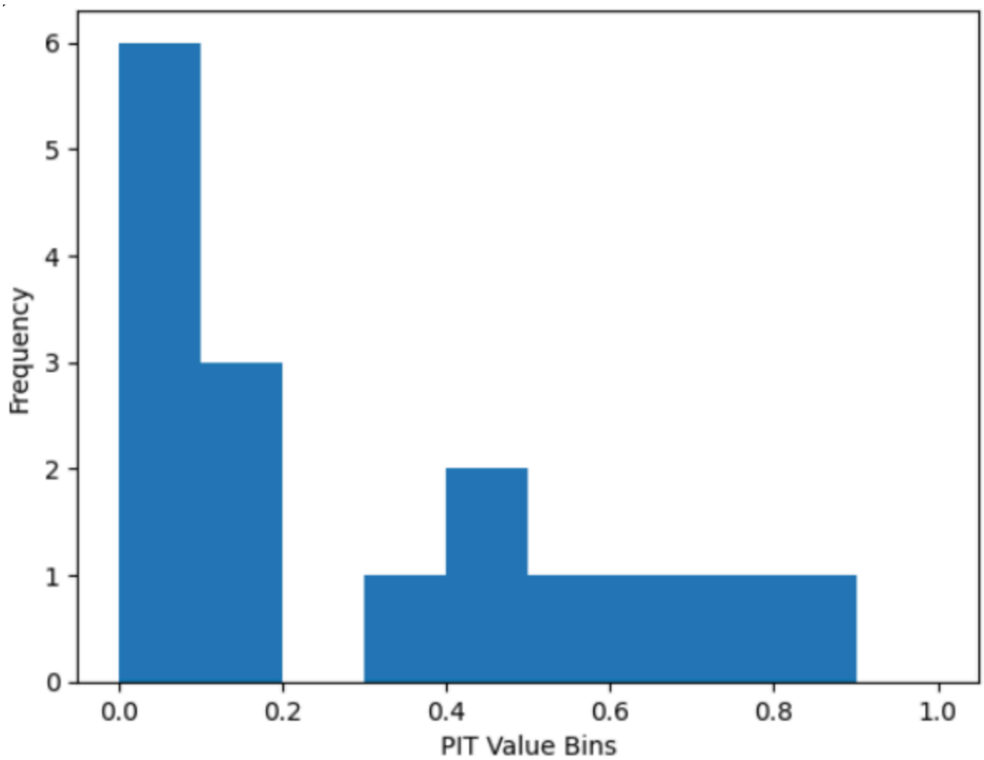
Note: The charts included in the question visually demonstrate the differences between what a well-calibrated vs. miscalibrated PIT distribution looks like.

Question 6 of 6

P2.T5.25.3.3

The risk team at Horizon Capital is reviewing their Value-at-Risk (VaR) model after a period of unexpected market volatility. To ensure the model is accurately predicting risks across all percentiles of potential losses, the team uses the Probability Integral Transform (PIT) distribution to evaluate the model's calibration.

During their analysis, the team observes patterns in the PIT distribution and discusses whether the model is well-calibrated or requires adjustments to better capture tail risks and market dynamics.



Which of the following observations would indicate that the VaR model may have been poorly calibrated?

- **A.** Well calibrated, the PIT distribution is uniform across [0,1], with no clustering in the tails or center.
- **B.** Poorly calibrated, The PIT distribution exhibits a concentration of values near 0, indicating frequent underestimation of extreme losses.

- **C.** Poorly calibrated, the PIT distribution shows slight deviations around the center but aligns well with a uniform distribution in the tails.

- **D.** PIT distribution has skewness close to 0 and kurtosis near 3, indicating balanced calibration of the model.

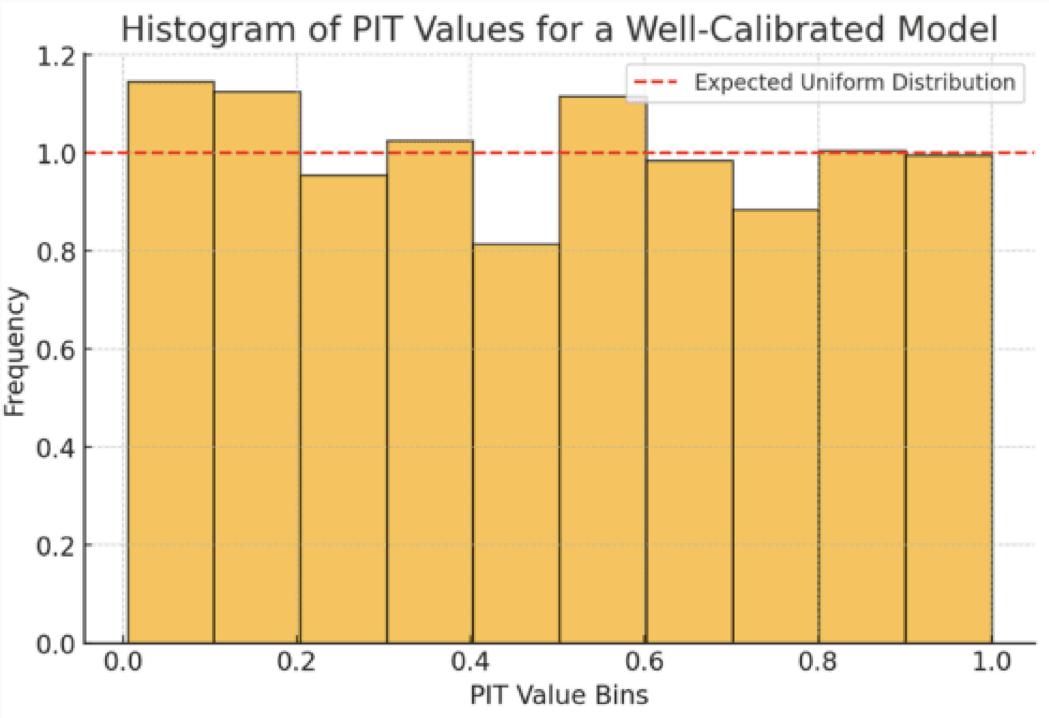
Explanation

B is CORRECT.

B. Poorly calibrated, the PIT distribution exhibits a concentration of values near 0, indicating frequent underestimation of extreme losses.

Clustering of PIT values close to 0 means the realized losses often fall into the worst tail (e.g. the 1% or 5% loss events) more frequently than the model predicts. That under-forecast of extreme negative outcomes is the hallmark of poor calibration in VaR models.

- **Regarding A,** although a uniform PIT distribution indicates a well-calibrated model, the distribution here is not uniform. A truly well-calibrated model *would* produce an approximately flat (uniform) PIT histogram (like below).



- **Regarding C,** slight deviations in the center with alignment in the tails generally indicate good calibration instead of poor.
- **Regarding D,** skewness near 0 and kurtosis close to 1.8 (the uniform distribution benchmark) suggest the model is balanced and well-calibrated. This is clearly not the case in our scenario.