

Exercises: Lesson 7.4 – Model Risk Governance and Validation
Module 7: The Compliance Problem

Prof. Dr. Joerg Osterrieder

Exercise 1: Model Tiering Assessment

A mid-sized European bank uses the following models. Assign each to Tier 1, 2, or 3, and justify your tier based on materiality and complexity.

#	Model	Use Case
A	IFRS 9 ECL model	Regulatory reporting: expected credit loss
B	Marketing propensity model	Predict likelihood of product purchase
C	AML transaction scoring	Flag suspicious transactions
D	Derivatives pricing (exotic)	Price illiquid structured products
E	Employee attrition predictor	HR workforce planning
F	Operational risk scorecard	Estimate operational losses

Tasks:

- Assign a tier (1, 2, or 3) to each model.
- For each Tier 1 model, state the required validation frequency and depth.
- Identify which dimension (materiality or complexity) drives the tier for models C and D.
- If model B's outputs were later used for credit limit decisions, how would its tier change? Why?

Exercise 2: VaR Backtesting Analysis

A bank's risk team reports the following VaR backtest results over 250 trading days at the 99% confidence level:

Quarter	Trading Days	Exceedances	Expected
Q1	63	1	0.63
Q2	62	0	0.62
Q3	63	2	0.63
Q4	62	5	0.62
Full Year	250	8	2.5

Tasks:

- In which Basel traffic light zone does the full-year result fall? What is the resulting capital multiplier?
- Are the exceedances evenly distributed or clustered? What does this pattern suggest?
- Using the Kupiec test, calculate the likelihood ratio statistic. Is the null hypothesis (model is correct) rejected at the 5% level? ($\chi_{0.05,1}^2 = 3.84$)
- What specific actions should the Model Risk Committee take based on these results?
- If the bank switches from parametric VaR to historical simulation, how might the exceedance pattern change?

Exercise 3: Champion-Challenger Test Design

A bank wants to replace its current logistic regression credit scorecard (champion) with a gradient boosted tree model (challenger). Both models have been trained on the same historical data.

Metric	Champion (LR)	Challenger (GBT)
Gini coefficient	0.68	0.75
KS statistic	0.42	0.49
PSI (6-month)	0.04	0.11
Avg. inference time	2 ms	45 ms
Explainability	Full (coefficients)	Partial (SHAP)

Tasks:

- Design a 12-month champion-challenger test plan: what data, what metrics, what success criteria?
- The challenger's PSI of 0.11 is below the 0.25 major-shift threshold but notably higher than the champion's 0.04. What does this suggest about stability?
- Calculate the percentage improvement in Gini. Is this improvement likely to be statistically significant across a typical loan portfolio of 500,000 accounts?
- Write a one-paragraph recommendation for the Model Risk Committee. Should they approve the challenger? Under what conditions?

Exercise 4: GenAI Use-Case Governance Assessment

A bank's innovation team proposes three GenAI use cases. For each, assess the governance requirements.

#	Use Case	Proposed Implementation
1	Summarize regulatory change documents	LLM reads new regulations, produces bullet-point summaries for compliance officers
2	Generate credit decision narratives	LLM writes the text explanation accompanying each credit decision letter
3	Chatbot for internal IT help desk	LLM answers employee questions about password resets and software access

Tasks:

- Assign a risk tier (1, 2, or 3) to each use case. Justify each assignment.
- For use case 2, identify three specific governance controls that must be in place before deployment.
- Describe two adversarial attack scenarios (prompt injection) for use case 1 and propose mitigations.
- Which use case has the lowest governance overhead? Explain why.
- Draft a one-page "AI use-case approval form" template with the key fields the Model Risk Committee would need to review.

Exercise 5: Audit Trail Design Challenge

A bank deploys a fraud detection model that processes 2 million transactions per day. Each model decision must be fully auditable.

Tasks:

- a List the minimum data fields that must be captured for each model decision (aim for 8–10 fields).
- b Calculate the approximate storage requirement for one year, assuming each decision record is 2 KB. Express in TB.
- c A regulator requests all model decisions for a specific customer over the past 3 years. Design the query interface: what search parameters, what response format, what maximum response time?
- d The bank uses a third-party cloud provider for storage. What contractual and technical controls are needed to ensure immutability and meet GDPR data residency requirements?
- e A GDPR “right to erasure” request conflicts with the 7-year audit retention requirement. How should the bank resolve this conflict? Cite the relevant legal basis.

Exercise 6: Sensitivity Analysis Interpretation

A model validation team tests a credit risk PD model by perturbing key inputs. The results are:

Input Variable	Base PD	PD at +1 SD	PD at -1 SD	PD at +2 SD
Debt-to-income ratio	3.2%	4.1%	2.5%	5.8%
Employment tenure (yrs)	3.2%	2.8%	3.5%	2.4%
Credit utilization	3.2%	4.5%	2.1%	12.7%
Number of inquiries	3.2%	3.4%	3.1%	3.6%
Account age (yrs)	3.2%	2.9%	3.6%	2.6%

Tasks:

- Which variable exhibits the most linear (proportional) sensitivity? Show your reasoning.
- Which variable shows evidence of a non-linear “cliff effect”? What is the evidence?
- Calculate the sensitivity ratio (PD change / input change in SD) for debt-to-income and credit utilization at +1 SD.
- Should the validator raise a finding for any of these results? Draft the finding with severity (low/medium/high).
- Propose a compensating control for the variable with the cliff effect.

Exercise 7: Model Risk Committee Decision

You are a member of the Model Risk Committee reviewing the following quarterly dashboard:

Metric	Q1	Q2	Q3	Threshold
Total models in inventory	2,340	2,380	2,415	–
Tier 1 models overdue	2 (4%)	5 (10%)	8 (16%)	<10%
Open validation findings	45	52	71	<60
Avg. finding age (days)	85	110	145	<120
Models without documentation	12	18	24	0
GenAI models in production	0	2	5	–

Tasks:

- Identify three metrics that are in breach of thresholds. For each, assess the trend direction.
- Draft a risk escalation memo (5–7 sentences) to the Board summarizing the key concerns.
- Propose three concrete actions the committee should mandate, with owners and deadlines.
- The 5 GenAI models were deployed without formal governance. What immediate steps are required?