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:

a. Assign a tier (1, 2, or 3) to each model.

b. For each Tier 1 model, state the required validation frequency and depth.

c. Identify which dimension (materiality or complexity) drives the tier for models C and D.

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:

a. In which Basel traffic light zone does the full-year result fall? What is the resulting capital multiplier?

b. Are the exceedances evenly distributed or clustered? What does this pattern suggest?

c. Using the Kupiec test, calculate the likelihood ratio statistic. Is the null hypothesis (model is correct) rejected at the 5% level? ($\chi^2_{0.05,1} = 3.84$)

d. What specific actions should the Model Risk Committee take based on these results?

e. 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:

a) Design a 12-month champion-challenger test plan: what data, what metrics, what success criteria?

b) 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?

c) Calculate the percentage improvement in Gini. Is this improvement likely to be statistically significant across a typical loan portfolio of 500,000 accounts?

d) 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:

a) Assign a risk tier (1, 2, or 3) to each use case. Justify each assignment.

b) For use case 2, identify three specific governance controls that must be in place before deployment.

c) Describe two adversarial attack scenarios (prompt injection) for use case 1 and propose mitigations.

d) Which use case has the lowest governance overhead? Explain why.

e) 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:

a) Which variable exhibits the most linear (proportional) sensitivity? Show your reasoning.

b) Which variable shows evidence of a non-linear "cliff effect"? What is the evidence?

c) Calculate the sensitivity ratio (PD change / input change in SD) for debt-to-income and credit utilization at +1 SD.

d) Should the validator raise a finding for any of these results? Draft the finding with severity (low/medium/high).

e) 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:

a) Identify three metrics that are in breach of thresholds. For each, assess the trend direction.

b) Draft a risk escalation memo (5–7 sentences) to the Board summarizing the key concerns.

c) Propose three concrete actions the committee should mandate, with owners and deadlines.

d) The 5 GenAI models were deployed without formal governance. What immediate steps are required?

Answer Key (Selected)

1. **Ex 1:** A=Tier 1 (regulatory capital), B=Tier 3 (low materiality), C=Tier 1 (AML regulatory), D=Tier 1 (high complexity + materiality), E=Tier 3, F=Tier 2. If B's output feeds credit limits, it becomes Tier 1 (materiality increases).
2. **Ex 2:** Yellow zone (5–9 exceedances); multiplier 1.50x for 8 exceedances. Clustered in Q4 (5 of 8), suggesting volatility regime change. Kupiec LR:
   $-2\ln \frac{(0.99)^{242}(0.01)^{8}}{(242/250)^{242}(8/250)^{8}} = 7.31 > 3.84$, reject null.
3. **Ex 3:** Gini improvement: $(0.75 - 0.68)/0.68 = 10.3\%$. PSI difference suggests GBT may be less stable across populations. Recommendation: extend test with focus on stability; approve only if PSI stabilizes below 0.10 over 12 months and explainability via SHAP is deemed sufficient for regulatory requirements.
4. **Ex 4:** Use case 1: Tier 2 (regulatory input but human review). Use case 2: Tier 1 (direct customer impact, GDPR Art. 22). Use case 3: Tier 3 (low risk, no financial decisions). Lowest governance: use case 3.
5. **Ex 5:** Storage: 2M txn/day $\times$ 365 days $\times$ 2 KB = 1.46 TB/year. GDPR vs. retention: AML regulations (AMLD5 Art. 40) provide legal basis for retaining transaction data despite erasure requests.
6. **Ex 6:** Credit utilization shows cliff effect: +1 SD gives +1.3 pp, but +2 SD gives +9.5 pp (non-linear). Sensitivity ratio: DTI = 0.9 pp/SD; utilization = 1.3 pp/SD at +1 SD but 4.75 pp/SD at +2 SD. Finding: HIGH severity for credit utilization non-linearity.
7. **Ex 7:** Breaches: overdue Tier 1 (16% vs 10%), open findings (71 vs 60), finding age (145 vs 120 days). All trending worse. GenAI models: immediate freeze of new deployments, retroactive validation within 90 days, governance framework before further GenAI expansion.