Validating Automated Transaction Monitoring Models: New Methods and Tools

Monday, April 3 | 4:50 PM

Moderator:
Robert L. Curry, CAMS, Executive Vice President, Chief Compliance Officer, Risk Management, KeyCorp

Speakers:
Rock T. Magleby-Lambert, Senior Manager, PNC Bank
Michael O'Malley, Director of Model Validation, PNC Bank
John Sabatini, Principal, Client Services, PwC
What Is a Model?

From the OCC 2011-12 / SR 11-7 definition:

“The term “model” refers to a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates.”

Key determinant of a model: uncertainty of the outputs

- Analytical tools explain data: “what happened?”
- Models make an inference about what might be true or might happen.
Why Are Automated Transaction Monitoring Tools Models?

• They make an inference about the likelihood that a pattern of transactions is “bad.”
  • Like other models, they are based on the bank’s historical data, which should be a good representation of the bank’s risk profile and efforts to preclude such behavior.

• They are different in how this inference is made.
  • They are often “rules based” as opposed to “equation based.”
  • … but they still meet the definition of a model.
Regulatory expectations

• Alignment of models with BSA / AML Risk Assessments
• Effective coverage of risks and red flags
• Ongoing monitoring and documentation of automated transaction monitoring system validation results
• Independent and skilled team to validate automated transaction monitoring systems in compliance with OCC 2011-12 / SR 11-7.
Whereas much of the focus of “AML validation” has been on the operational aspects of the alert generation and review processes, an additional focus of model risk validation needs to be on how the models / scenarios are built and whether they can be trusted to ensure adequate differentiation of good vs. bad activity.
Start with a Risk Assessment and Coverage Gap Analysis

- Risk coverage analysis aims to ensure that inherent and residual risks are being effectively monitored and managed through the controls in place
- An effective validation of risk coverage should take into account:
  - Review of industry risk factors & BSA / AML Risk Assessment
  - Automated BSA / AML risk factors and “red flags”
  - Typology assessment and scenario coverage (active/inactive inventory)
  - Regulatory guidance and expectation

### Risk Assessment Heat Map

<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>Risk</th>
<th>Scn1</th>
<th>Scn2</th>
<th>Scn n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Factor 1</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>x</td>
</tr>
<tr>
<td>Risk Factor 2</td>
<td>M</td>
<td>A</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Risk Factor 3</td>
<td>L</td>
<td>I</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Risk Factor 4</td>
<td>H</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Risk Factor 5</td>
<td>H</td>
<td>A</td>
<td>I</td>
<td>x</td>
</tr>
<tr>
<td>Risk Factor n</td>
<td>M</td>
<td>x</td>
<td>A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- A Adequate
- I Inadequate
- x No Coverage

**Review Gaps**

Review identified gaps with the bank. During this process, review findings with the bank to determine a course of action appropriate for the bank. Items to be reviewed are: data needed to run the reports and the tactical vs. strategic plan development.

**Develop Plan**

Software Core Platform
Planning: Understand and Prioritize Higher Risks

- Group and classify models in various surveillance categories and prioritize them based on a risk weighting methodology to identify the most critical.

- Determine the effectiveness of models individually and within the given families by reviewing the productivity metrics.

- Metrics are an important measure of the validation process.

- Review the underlying technical and functional logic of the model against the business requirement and/or the regulatory mandate requiring the surveillance.

- Scenario logic can be developed independently to validate the output with that of the model.
Planning: Have specific performance metrics. Then tune and monitor to achieve them.

Metrics to Consider:

- **Recall** – the number of stars you catch divided by the total number of stars in the graph. (It’s all about the stars.)
- **Precision** – the number of stars you catch divided by the number of stars and circles you catch. (It’s all about what you catch.)
- **False negatives** – the number of stars you don’t catch. (Note the inverse relationship with Recall.)

*Increasing Recall often comes at the expense of decreasing Precision!*

Data: Ensure Appropriate Inputs for Both Tuning and Production

Ensure data are sufficient for both tuning and to then to ensure complete and effective alert generation after scenarios are implemented.

Ongoing monitoring through process verification and benchmarking is a core element of validation. The assessment involves an analysis of the different data inputs from source systems through ETL, staging areas and surveillance.

**Data Sources**
- Customers, Accounts & Transactions
- ETL Mappings, Logic Exception Logs
- Alerts, Cases data

**Data Validation & Testing**
- Load data for testing
- 6-12 months of Data
- Ongoing Monitoring of Data

**Validation and Testing Components**
- Ongoing Monitoring Controls
- Data Flow Assessment
- ETL & Data Filters
- Data Quality, Completeness and Integrity
- Data Mapping and Transformation
- Models/Reports Selection & Testing
- Risk Coverage
- Productivity Testing
One size does not fit all. Differentiating how alerts are generated for different exposures (“segmentation”) is a key to hitting performance metrics while balancing the costs of false positives and false negatives.

By grouping customers or accounts with similar transaction activity (e.g., products, services, transactions, and geographic locations), alerting schemes can be tailored to efficiently differentiate risky transactions.

**Key Consideration:**

- a) Validate data inputs and relevant business rules
- b) Conduct statistical clustering analysis around centrality, homogeneity, stability, etc.
- c) Evaluate cluster effectiveness through quantitative analysis to verify that population distribution across segments and segment transaction activity aligns with business understanding
- d) Document findings of the analysis and compare against current segmentation model outputs to evaluate current segments.
Tuning: Thresholds must differentiate risk

- Many “estimators” can be used to differentiate the likelihood of good vs bad transactions:
  - Rules
  - Judgmental scores
  - Regressions
  - Pattern analysis, etc.

- Regardless, “tuning” must involve a trade-off between the respective costs of false positives (precision) and false negatives (measured by recall).
  - Thresholds can then be set given a tolerance for this trade-off.

- No matter what approach you take, having explicit performance metrics should be used in both the tuning and monitoring components of model risk validation.
Tuning: Using Performance Metrics

Risk is Rank Ordered by the Amount of Transactions in this example.

- Precision increases with larger transactions: larger transactions are more likely to be interesting.
- But where do we set our threshold to alert?
  
  A: Alert all interesting activity but lots of false positives (low precision)
  B: Some false negatives (lower recall), but fewer false positives (higher precision)
  C: More false negatives (lowest recall), but even fewer false positives (highest precision)
Production Monitoring: Above and Below the Line Testing

• Let’s say you chose “B” in the previous slide.

• Are you achieving the desired precision above the line?

• Are you testing that the rate below the line is still within tolerance (i.e. recall)?

• Does your monitoring suggest that the risk has shifted or that your thresholds are suboptimal?
  • If so, consider re-tuning.

Are we achieving 23% precision above the line and roughly 6% below the line?
• If not, why not?
• If we get more than 23% AtL, does that mean there’s more risk just above the threshold? If so, does that imply there’s more risk just below that we’re not alerting?
Model Output: Threshold Validation

Threshold optimization focuses on validating the initial threshold values that have been set across segments based on statistical analysis, alert coverage and historical alert productivity. The goal is to increase the productivity and reduce the false positives generated by business rules.

**Key Consideration:**

a) Historical SAR information

b) Conduct independent baseline threshold and outlier analysis to calculate and analyze outlier percentile amounts between 95th & 99.9th percentile.

c) Identify and investigate potential productive reports below the line, or non-productive reports above the line

d) Use the results from the ATL/BTL testing to evaluate current threshold settings

* Required in instances where business rules have multiple alerting criteria or sub-rules
Model Output: Risk Scoring

Alert Risk Scoring focuses on developing a risk-based approach to the alert review process. The process assigns numerical values (scores) to alerts based on statistical models and business rules to determine the priority and relative risk level of an alert.

Validating a risk scoring model involves independent testing and recreation of model logic using statistical techniques to evaluate results.

Key Consideration:

a) Validated requirements of the scoring models against business objectives
b) Qualitative review of attributes selected for model development
c) Statistical Verification: Use statistical techniques (e.g. Linear Regression, Logistic Regression, Regression Trees etc.) to verify selected attributes used for developing scoring models
d) Evaluation: Leverage statistical techniques (such as ROC curves) to verify model outputs and perform outcome analysis (effect on output as input changes) to check for sensitivity of the model.

* Required in instances where business rules have multiple alerting criteria or sub-rules
A Sign of the Times: NYDFS 504 and Model Validation

Applicable NYDFS Part 504 requirements

• **504.3(a).5:** End-to-end, pre-and post-implementation testing of the Transaction Monitoring Program, including, as relevant, a review of governance, data mapping, transaction coding, detection scenario logic, model validation, data input and Program output

• **504.3(c).7:** Qualified personnel or outside consultant responsible for the design, planning, implementation, operation, testing, validation, and on-going analysis, of the Transaction Monitoring, including automated systems if applicable, as well as case management, review and decision making with respect to generated alerts and potential filings.

Key Considerations

<table>
<thead>
<tr>
<th>Model Validation knowledge</th>
<th>How to enhance your BSA/AML Program elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adopt an aggressive hiring strategy to hire individuals from the industry or from regulatory agencies with significant quantitative and statistical knowledge to be able to analyze models</td>
<td></td>
</tr>
</tbody>
</table>

| Documenting Roles and Responsibilities | Establish a centralized model risk management group (MRMG) that oversees the model validation of all AML models |

| Dynamic Regulatory Landscape | Ensure that the in-house MRMG is provided with adequate and appropriate trainings and exposure to regulatory conferences in order for them to keep up with the dynamic regulatory landscape and constantly changing requirements from the regulators |
Key Takeaways

• Understand regulatory expectations for model governance and how they relate to your specific program, practice and risk assessments (OCC 2011=12 / SR 11-7).

• Model validation needs to consider data, risk and coverage assessments.

• Tuning models (and non-models) should lead to increased effectiveness (data inputs, statistical analysis, typology development).

• Ensure your program has resources that have quantitative and statistical background to challenge / enhance your journey to continuously strengthen transaction monitoring surveillance systems.