# **Ongoing Performance Monitoring (OPM) Plan: Instructions**

The objective of the OPM plan is to develop a systemic approach to monitor, evaluate, and optimize the ongoing performance of models to ensure accuracy, reliability, and relevance of the model over time.

## **The OPM Plan**

The OPM plan needs to contain the following:

1. Metrics for Performance Measurement.
2. Thresholds (three categories: Green, Yellow, and Red).
3. Threshold Breach Action Plans.
4. A model status summary.
5. The OPM report frequency.
6. A Sample of the OPM Report.

## **1.1 The Model Performance Metrics**

In general metrics fall into a few groups:

* **Data Metrics** track any changes in the model input data that are used for the model’s production process, such as:
* Data Trend Analysis for outliers and missing values that could affect model’s performance.
* Population Stability Analysis to ensure the portfolio population characteristics are not materially shifted over time.
* Model Input Variable Characteristics Analysis.
* **Performance Metrics** track how well the model is performing at its basic task, and whether current model performance falls within the expected range. These metrics’ selection needs to be model specific with strong justification/rationales to support your selection of each metric such as:
* Academic references
* Historical performance
* Industry best practices
* **Diagnostic Metrics** test if the model has developed problems such as deviations from model assumptions. Since not all models make explicit assumptions about how the data behaves, this section, too, may not always be present in the OPM report. Again, the choice of diagnostic metrics must be supported by appropriate references.

The metrics should be clearly defined in the OPM Plan.

## **1.2 The Thresholds for Metrics**

The metric thresholds must be set so three categories are defined:

* **Green** – The model is performing as expected. No further action/analysis needed.
* **Yellow** – The model is performing with some early warning concerns, but not fatal.
* **Red** – The model is not performing as expected.

Crossing the first threshold puts the metric in Condition Yellow, crossing the second threshold puts the metric in Condition Red. If a metric is in Condition Yellow the first time, an explanation should be provided. If a metric is in Condition Yellow two reports in a row, at minimum a root cause analysis is required.

Condition Red should immediately trigger additional analysis to identify the drivers of the performance deterioration. The outcome of that analysis may lead to model changes, to temporary judgmental overlays or even to a completely new model.

Both thresholds should be defensible, either on statistical grounds or through academic/industry references. Thresholds that are set too high will lead to ineffective performance monitoring, thresholds that are set too low will lead to “false positives” and unnecessary work. All metrics and thresholds should have a baseline analysis using the Bank’s portfolio historical performance data. **All metrics and thresholds must be approved by MRM.**

## **1.3 The OPM Threshold Breach Action Plans**

For each metric’s threshold breach, an action plan should be well-defined and documented. These action plans may include:

* Perform a deeper analysis on the input data.
* Examine whether any parameters need to be recalibrated.
* Compare results with the results from a benchmark model, if available.

If no cause for the model’s performance deterioration can be identified, consult with MRM.

## **1.4 The Model OPM Status Summary**

There is no “rule” for how many metrics need to be monitored. This needs to be determined by the model owners. However, MRM does ask that a “summary” status is reported. Metrics can be grouped (e.g. data metrics, performance metrics, diagnostic metrics) and per group, a summary status is defined as follows:

* If any of the model metrics in the group are in Condition Red, then the Summary Status for that group is Red.
* If not, then if any of the model metrics are in Condition Yellow, then the Summary Status for that group is Yellow.
* If not, then the Summary Status for that group is Green.

An overall model status should also be reported by taking the summaries by group and applying the same rules to come up with a summary status for the model as a whole.

## **1.5 Reporting Frequency**

The OPM Report should be issued at least quarterly. For models which are only used annually, consult with MRM.

## **Example of an OPM Report**

**Ongoing Performance Monitoring Report**

|  |  |
| --- | --- |
| Name of Model | PD |
| Model ID | 42 |
| Model Owner | XXX XXXXX |
| Report Date | Q2 2025 |

**Overall Model Status**

|  |  |
| --- | --- |
|  | Performance Metrics in Condition Yellow |

**Detailed Metrics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Metric** | Current | Value |  |  |  |
| **Performance Metrics** | | | | | |
| Mean Absolute Percentage Error |  | 4% | <5% | 5-8% | >8% |
| Absolute Percentage Error |  | 7% | <6% | 6-9% | >9% |
| **Diagnostic Metrics** | | | | | |
| Hosmer-Lemeshow P-value |  | 12% | >5% | 1-5% | <1% |
| McFadden R-squared |  | 24% | >15% | 5-15% | <5% |

**Comments**

An upturn in the most recent APE caused it to pass the Yellow threshold. The initial root cause analysis suggests the recent GDP growth numbers may be producing forecasts that are too optimistic. If this continues, recalibrating the model may be necessary. The diagnostic metrics are all satisfactory. Note that the MAPE is still Green due to the smoothing effect of averaging.

**Definitions**

### Metric 1: Mean Absolute Percentage Error

Definition: Percentual one-month ahead forecast error of the model, averaged over the last quarter

Threshold 1: 5%, based on 90th percentile of historical model performance

Action Plan 1: Review model inputs for unusual values and their impact

Threshold 2: 8%, based on 95th percentile of historical model performance

Action Plan 2: Full root cause analysis

Etc.

## **Appendix: Some Ideas for Metrics**

|  |  |
| --- | --- |
| **Possible Approaches** | |
| 1. | Outcome analysis by comparing with actual outcomes |
| 2. | Benchmarking |
| 3. | Back-testing (out-of-sample, out-of-time, and/or value-at-risk (VaR) analysis) |
| 4. | Model sensitivity, boundary, and stress analysis |
| 5. | Population stability analysis |
| 6. | Model input variable characteristic analysis |
| 7. | Scenario analysis |
| 8. | Model input data (internal and external) accuracy, completeness, and consistency analysis |
| 9. | Tests implemented during the model development |