Auditing the AML/CTF Transaction Monitoring System

Jon W Harvey CAMS-AUDIT, CAMS, CGSS

April 2020

Table of Contents

Table of Figures

Executive Summary

This paper considers an approach to performing a technical audit of the Transaction Monitoring (TM) system. Whereas the role of the TM system in the AML control framework is well understood by financial crime professionals, less is known about the configuration considerations and the options that should be reviewed to provide assurance in respect of the effectiveness and efficiency of the system.

Having a backlog of alerts is not unusual, but how often does the auditor evaluate the composition of the alert queue or for that matter reconcile the TM system with the key risks identified by the risk assessment or assess the efficiency of detection rules or the root cause of the ubiquitous false positives?

This paper is focused on obtaining assurance that the TM system is appropriately configured. It does not discuss in detail the closely related processes of risk assessment, alert workflow and investigations, disposition of alerts, or investigation operations.

Introduction

The obligation to perform transaction monitoring is defined in AML/CTF regulation worldwide. Regardless of jurisdiction, the objective remains the same, which is to monitor a client's transactions and identify those that are inconsistent with the bank's knowledge of the client, the client relationship, and the expected use of the financial product.

In addition to regulatory obligations to assess AML/CTF risk and establish an appropriate TM system, expectations for management of the TM system are defined in several useful reference texts.

- Department of Financial Services Superintendent's Regulations Part 504 BANKING DIVISION TRANSACTION MONITORING AND FILTERING PROGRAM REQUIREMENTS AND CERTIFICATIONS - https://www.dfs.ny.gov/docs/legal/regulations/adoptions/dfsp504t.pdf
- Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency SUPERVISORY GUIDANCE ON MODEL RISK MANAGEMENT OCC 2011-12 https://www.occ.treas.gov/news-issuances/bulletins/2011/bulletin-2011-12a.pdf
- UK Financial Conduct Authority(FCA) Financial Crime Guide: A firm's guide to countering financial crime risks (FCG) https://www.handbook.fca.org.uk/handbook/FCG/2/?view=chapter
- Post-event transaction monitoring process for banks Guidance De Nederlandsche Bank https://www.toezicht.dnb.nl/en/binaries/51-236846.pdf

In its report, "Guidance for a Risk-Based approach The Banking Sector" (FATF, 2014, p. 21), the Financial Action Task Force[1] states that "*Monitoring should be carried out on a continuous basis or triggered by specific transactions. It could also be used to compare a customer's activity with that of a peer group." Further, it states that "Banks should adjust the extent and depth of monitoring in line with their institutional risk assessment and individual customer risk profile*s."

[1] Financial Action Task Force (FATF) is an inter-governmental body established in 1989 by the ministers of its member jurisdictions. The objectives of the FATF are to set standards and promote effective implementation of legal, regulatory, and operational measures for combating money laundering, terrorist financing, and other related threats to the integrity of the international financial system.

TM systems are used to evaluate client transactions and identify activity that is unusual for the client type. This simple statement contains two very important concepts that are key to the correct configuration of the TM system. Segmentation of clients into homogeneous groups or segments enables us to understand what constitutes the usual behaviour for clients in the same segment. A transaction by one type of client could be routine, whereas for another type it is suspicious and should be investigated. Using client attributes to perform segmentation makes it lot easier to identify unusual or anomalous behaviour.

The UK Financial Services Authority (now Financial Conduct Authority) states in its report, "Automated Anti-Money Laundering Transaction Monitoring Systems" (FSA, 2007, p. 1), that TM systems[2] use profiling and/or rules-based monitoring methods to identify unusual patterns of customer activity by applying statistical modelling techniques, or rules-based monitoring, which compares customer activity to fixed pre-set thresholds or patterns to determine if it is unusual. Rules based monitoring is used in TM systems, such as the market abuse monitoring where rules and thresholds are known.

If we fail to calibrate the monitoring system to be sensitive to the difference between expected transactions and those that are anomalous for the client type, we will have an inefficient system that generates false positives, and, more importantly, may fail to identify potential money laundering in an acceptable timeframe.

The approach outlined in this paper considers the audit of the TM system through its lifecycle of implementation and maintenance, through to consideration of the source data that is fed into the TM system, the data that is provided to the investigator for the timely, efficient, and accurate disposition of alerts.

Approach

The TM lifecycle starts with identifying appropriate data sources. Data is subject to cleaning, enrichment, and augmentation before being assembled ready to be fed into a process that segments clients, evaluates the client transactions against appropriate risk scenarios, and creates alert events for transactions that meet or exceed the thresholds specified for each scenario with alert events going into queues to be investigated by analysts. This is intentionally a very simple explanation of the lifecycle; the sections that follow expand on the processes and discuss what takes place at each stage, as well as where the auditor can obtain assurance for the TM system effectiveness.

[2] https://www.fca.org.uk/publication/archive/fsa-aml-systems.pdf

The approach described below considers the TM system as comprising the following stages:

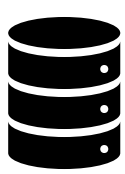
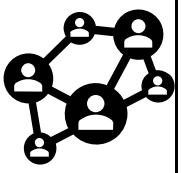
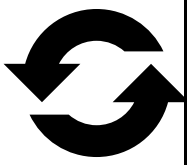
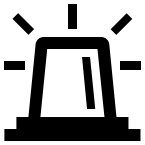
Source Data	Data Cleaning	Segmentation	Transaction Monitoring	Alerts	Investigate
					
Client data/KYC Transactions Reference lists	Correction Enrichment Consolidation	Clients Clustered by Key Attributes	Transactions Screened Against Scenarios Using Thresholds	Transactions "Scoring Enough Points" Generate Alerts	Analysts Investigate Alerts

Figure 1: The Transaction Monitoring Process

Source Data

The feed of transaction data to the TM system is produced by combining data from various sources typically originating from disparate systems. Data sources are combined to provide the context of the transaction so that it can be evaluated.

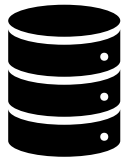
Source Data	Risks	Controls
	• Incorrect configuration of the TM system may result in transactions that shoud be screened, not being subjected to screening, and therefore, failure to detect anomolous activity. • TM system does not function correctly if it is not provided with reliable, accurate data. • TM system fails to identify transactions that should be investigated. The organisation may fail to prevent/report actual or suspected money laundering/terror financing.	• System testing shoud be performed to ensure that all transactions, regardless of origination, are reconciled to monitoring controls. • Data lineage controls to understand origin, completeness, format specification, reliability of data, and its transformation throughout the transaction life cycle. • Periodic review of data sources; their security and integrity. • Data governance controls to ensure data ownership and custodianship arrangements. • Formal data management and change control to protect data and its use.

Figure 2: Source Data Risk and Controls

The TM system needs information that resides in various files, such as:

- client file information including the KYC data, client identifier, date of account opening, client risk score, recent alert data, the results of due diligence, etc.;
- transaction information, such as amounts involved, beneficiary information, geography, delivery channel used, etc.;
- reference sources such as geographic risk data, sanction lists, PEP, and watchlists, etc.

When considering the source data, it is important to have regard for the data lineage. Data lineage refers to the origin of the data; how it is sourced and transformed throughout the lifecycle. By considering the data lineage, we can ensure that data is accurate, relevant, and from a reliable source.

From an audit perspective, if we know that there are issues with data quality, it is important to consider how the TM system deals with cleaning the data and making sure the data required at scenario level is available, accurate, and complete. If data used to screen a

transaction against a scenario is omitted or not reliable, then the scenario cannot be reliably applied.

Whilst auditing source data, it is important to confirm that in addition to the availability of reliable data, all transactions are subject to monitoring. Errors can occur in systems configuration, such as badly configured routing tables, incorrect parameters for transactions originating from different geographies, incomplete systems mapping, or products that are not subject to appropriate scenarios.

In 2018, a European banking group (Santander Consumer Bank , 2019, p. 9)[3] identified that its TM system had not been configured correctly, and an excess of 1.6 million transactions performed by an excess of 300,000 clients over four years had not been subject to monitoring. Subsequent review of the transactions identified a number of the transactions were performed by 32 high-risk clients.

The banking group self-disclosed to the Norwegian FSA and undertook a look-back review, stating that it did not identify transactions that should be disclosed. This weakness resulted in several transactions not being monitored in line with the requirements of the Money Laundering Act (Hvitvaskingsloven). The bank was given a financial penalty of 9 million NOK for breach of the money laundering act.

Data Cleaning

Data cleaning is essentially the process of confirming that the data required to apply a transaction monitoring scenario to a transaction (the scenario parameters) is accurate, reliable, and timely. This control applies to all data, such as ensuring that KYC files are compliant, complete, and maintained. In addition to KYC and transaction details, the principles of data cleaning need to apply to data, such as the availability of the organisation's high-risk geography information and details of which delivery channel was used to originate the transaction, both of which would have a bearing on determining the transaction score.

Data Cleaning	Risks	Controls
	• TM system does not function correctly if it is not provided with reliable, accurate data. • TM system screening may fail to apply appropriate scenarios and thresholds resulting in failure to alert transactions that should be investigated. • TM system may fail to identify correct attributes of a scenario where transaction codes are not consistently applied. • TM system may fail to apply sufficient weight to a repeat alert if previous alerts are not identified as associated with current alert events. • The organisation may fail to prevent/report actual or suspected money laundering/terror financing. • Client industry type risk is not correctly assessed in consequence of failure to evalute the industry code correctly.	• Data lineage controls to understand origin, completeness, reliability of data, and its transformation throughout the transaction life cycle. • Data governance controls to identify ownership, data custodian and change control. • KYC procedures that define the due process for onboarding clients and maintaining client files. • Training materials and sessions that explain the due process for obtaining client data. • Management information, including operational data, data completeness, transaction repairs, and overides. • Testing of data completeness—reconciliation with conversion look-up tables, reconciliation of data fields required for scenarios, etc. • Compliance monitoring reviews with appropriate sampling and escaltion of exceptions.

[3] https://www.santanderconsumer.no/globalassets/om-oss/investor-relations/financial-reports/santander-consumer-bank-nordic-group-_q2-2019-report.pdf

		• Review of TM system exception audit reports, such as scenarios with bad or missing thresholds. • Review of records of action taken in response to management information and system reports. • Client industry type codes are defined in data governance policy. Mapping is applied where disparate conventions are in use.

Figure 3: Data Cleaning Risk and Control

Combining data sources inevitably results in duplicate data, data inconsistencies, missing data, and the same data items recorded in different formats. This is where data cleaning must take place and possibly where the audit starts to become more technical.

Scenarios may not be effective in detecting unusual transaction activity if they are not provided with accurate, reliable, and timely data. There is a risk that an ineffective scenario will fail to detect unusual activity that may be associated with money laundering or terror financing. TM systems will often produce an exception report that lists instances where a scenario was applied but key threshold parameter data was omitted, or was not within permissible range or format. This is a key report to identify and review; the auditor should verify the system is configured to produce such a report, review how the TM system performs with missing variables, which reports are generated, actions taken, and the escalation process.

Data cleaning is not limited to ensuring that data is present and in the correct format, it can include augmentation and code translation. Disparate systems may use different conventions to convey the same type of information, for example, if corporate clients have an industry-type code (SIC, NAICS, GICS, MSCI, etc.), then we may want to convert to a standardised code using a look-up table that can then be used by the TM system scenarios to evaluate inherent client risk.

A vital step in data cleaning is normalisation of the data. This is typically performed by the database administrators and would be performed for the normal functioning of the firm's application systems. When data has been normalised, the relationships between the data elements are understood and mapped. This is key to understanding client-to-account relationships, such as which accounts are associated with which clients, which clients may have multiple accounts, which accounts may be linked, and all accounts must have an owner, etc. The TM system cannot accurately assess transaction risk if the client relationships and associated accounts are not correctly mapped.

Note: If there are audit findings in respect of KYC files, it follows that data required to apply a transaction monitoring scenario may not be reliable, and we cannot have assurance that the transaction monitoring system is effective.

Segmentation

If we group clients by the business line that they have a relationship with, we obtain very broad groups. For example, if all retail banking customers are in the same group, we are potentially comparing students with high net worth individuals or even small businesses. Clearly, they have very little in common in terms of their financial behaviour.

Segmenting clients according to the business line that they have a relationship with may not be appropriate. An example would be a client in the retail segment that runs a small business, makes payments, and pays limited payroll from their personal account.

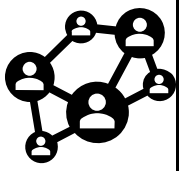
Segmentation	Risks	Controls
	• Incorrect segmentation of the client base undermines the risk-based approach by not monitoring clients for the risks that they pose or are exposed to. • Incorrect segmentation may result in failure to screen against appropriate scenarios that have been calibrated with appropriate thresholds for clients in the respective segments. • Incorrect segmentation may result in a client not being subject to the appropriate scenarios and thresholds potentially resulting in failure to detect, investigate, prevent, and report acts of known or suspected money laundering/terror financing. • Clients allocated to an incorrect segment will generate false positives, resulting in efficient operations, potential backlogs, and unnecesary cost. • Incorrect segmentation may result in the changes in client behaviour being undetected and a KYC review not being performed. Failure to maintain KYC in response to trigger events propagates errors in subsequent monitoring. • Incorrect segmentation may result in failure to detect connected/repeated event alerts. • A client may not be subject to appropriate monitoring in consequence of dynamic resegmentation.	• Appropriate segmentation that evaluates client attributes and transaction behaviour to segment clients • Segmentation should be subject to periodic review to ensure that it remains appropriate for the clients served, products and services, geography, and delivery channels used. • Maintenance of adequate documentation for the approach to segmentation, including data lineage, assumptions, and where segmentation should be reperformed (excessive polarization, orphaned clients, results from below the line testing), etc. • MI reports should be produced to report client segmentation, changes in segment composition, concentration/polarization, changes in risk, risk/alert distribution, event alerts distribution by segment, etc. • Client segment, scenarios, thresholds applied, and the relative "weight" or score of an alert should be subject to periodic review by business risk for approval. • Segmentation testing is performed using use cases or proxy clients to ensure appropriate scenario and thresholds are applied. Narrative should be provided to describe testing performed and results. • TM systems that perform dynamic segmentation typically produce a report; this should be subject to review and signoff.

Figure 4: Segmentation Risk and Control

If clients in a population are segmented according to those with similar attributes, it follows that we would have homogenous segments and should observe similar client behaviour or use of the products and services, and therefore, unusual activity is easier to identify.

The selection of attributes to use for segmentation is important and should be periodically reviewed. In practice, the precise attribute values are not used for segmentation; instead, the clients are segmented so that they are in the closest relevant segment. This approach ensures that we have a reasonable number of client segments and retain the homogeneity we require to identify anomalous behaviour.

Poor segmentation results in thresholds being set to values for broad populations with the associated wide range of behaviours as opposed to setting them specifically for the narrower range of behaviour of similar clients in the same segment. Thresholds can be more accurately set, and outlier behaviour more easily identified when segmentation is maintained.

A common issue with segmentation is that it reflects a point in time; therefore, thresholds that are set to identify unusual transactions within a segment will, over time, become less effective, as what we consider usual behaviour will evolve, and perhaps the distinction between usual and unusual becomes blurred, resulting in false positives. Another consequence of not maintaining the segmentation is that whilst we are monitoring for

transactions that were deemed unusual at the time of original threshold configuration, we are potentially failing to implement scenarios and thresholds to detect new and emerging risk and anomalous client behaviour.

Client behaviour will naturally evolve over time, but this is typically experienced across all clients in the same segment. Having said this, if client behaviour within a segment deviates over time, the TM system must continue to be able to detect outliers as exceptional to the peers within the segment.

Dynamic segmentation is used by some systems to reassign clients into different segments where client attributes, such as monthly funding, change over a specified timeframe. The treatment of dynamic segmentation, such as the logic to reassign to a different segment and threshold values, should be reviewed by the auditor to ensure that reassignment to another segment is appropriate. TM systems that perform dynamic segmentation typically produce a report; this should be subject to review and signoff.

Aside from generating an alert, the TM system may provide management reports, and these can include:

- clients that have changed their behaviour in such a way that the account should be reviewed, perhaps by way of a KYC refresh;
- clients that are generating multiple linked alerts;
- clients that have been, or should be, moved between segments; and
- clients that have performed transactions that have previously generated alerts (although not necessarily securing sufficient score to be investigated) and have changed segment.

There are many ways to approach segmentation, including simple intuitive approaches, such as which primary business line the client has a relationship with or use of marketing categorisation. Segmentation at this level is unlikely to be sufficiently sensitive to identify anomalous behaviour, as the segment is so large that the diverse range of activity would constitute normal activity, and therefore, outlier activity would have to be very exceptional to be identified. For example, if we consider a retail banking client population if we used a relationship with the retail business to allocate a client to retail sector, then we'd be monitoring the student accounts alongside the clients that had retired and the potential higher net worth clients.

Attributes that help provide greater segmentation include:

- Dominant transaction types used and frequency
- Client occupation or industry sector codes
- Size of client business
- Account turnover
- Maximum monthly balances, average amount paid away per month
- Use of international payments facilities
- Other data from KYC records such as PEP status, geography, or dominant transaction channel used

If we analyse the total transaction data by client, transaction type, frequency, and value, and eliminate low-risk transaction types (such as balance enquiries), then this would provide a distribution that can be used to segment the clients.

For example, if we group clients according to the monthly volumes and values of incoming wire transfers, then we can apply a scenario for accounts that received exceptional volumes and values when considered alongside peers.

Regardless of the approach and attributes chosen to inform segmentation, it is clear that segmentation will only be useful if the attribute data exists and is reliable; this is the purpose of the data clean-up phase in transforming the source data so that it can be used for segmentation.

Machine Learning Based Segmentation

In practice, most organisations use a machine learning technique named "K-means" to perform the process.

The "K" in K-means refers to the number of clusters that should be used to group together data that have similar attributes. For example, if we had a population of customers, and we choose to split them into two clusters, then we could use their registered address to allocate them to a cluster for customers in the North or the South. We could use more clusters, for example, to group together people that have retired, people that are students, people that are too young to work, and those that have a full-time job. The K-means technique will allocate data to the cluster determined to have the greatest similarity.

Each cluster has a value within the cluster called a centroid. The K-means algorithm measures the distance[4] between centroids and data attributes (in this case, the client attributes) and allocates the client (using the selected attributes) to the cluster that has the shortest distance to the centroid; in other words, is the closest fit. This is a very simplistic explanation of what takes place, and in practice, the centroid is moved within the cluster to the calculated mean, and the process repeated, so all clients are in the most appropriate cluster.

Use of the K-means technique to produce the clusters is computationally intensive; each additional attribute that is considered significantly increases effort.

K-means is an iterative technique until clusters are stable. Clusters are considered stable when clients are no longer being moved between clusters, as they have the best fit (shortest distance). On completion, distribution of clients across the segments should be reviewed to ensure a reasonable distribution. For example, if we segment a retail banking client base, we would expect to see reasonable separation of retail client types (students separated from clients that are employed with an average salary, clients that are higher net worth separated from retirees, etc.).

[4] The technique used in K-means to measure distance is the Euclidean Distance; this accurately measures the degree of fit and variance between the dataset (the client attributes) and the centroid within each cluster (K).

The clusters produced by the K-means process are our customer segments.

When auditing the approach to segmentation, it is important to consider the initial setting for K. If K is too large, then the process is slow to perform, and if it is too small, we don't get the required degree of homogeneity in our segmentation. If incorrect attributes are used for the clustering, the segmentation may not be homogeneous for the purpose of applying a scenario and appropriate thresholds. The general rule for setting K is the smaller the value, the more reliable the algorithm, the larger the value of K the more computationally demanding.

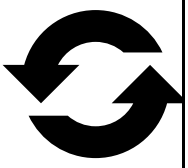
In terms of risk, we must consider if additional segmentation enables us to be more targeted with application of scenarios and thresholds, or if there is a diminishing return with greater administrative workload.

Whichever approach to segmentation is adopted, it should be fully documented, clearly identifying the data sources, data cleaning, and enrichment that is performed in order for the segmentation to take place. The description of the approach should include the period of time for which data is viable, and, in the case of static segmentation, the frequency that segmentation should be reperformed.

Monitoring

The monitoring stage is the heart of the process. This is where transactions performed by the client are screened against relevant scenarios that have been calibrated for expected client behaviour within a client segment.

Clients are associated with a segment, because they have similar attributes and behaviours. It follows that in addition to common scenarios that all clients would be subject to, scenarios specific to the segment and client risk profile should be used to monitor the client's transaction.

Monitoring	Risks	Controls
	• The organisation fails to detect actual or suspected money laundering/terror financng in consequence of the TM system not being configured to monitor for the risks identified in the risk assessment. • Incorrect segmentation may result in the transaction not being subject to the appropriate scenarios and/or thresholds. • Limited use of scenarios that evaluate for a range of risk factors may bias scenarios selection and reduce the significance/weight associated with events. • Incorrect segmentation may result in clients not being subject to monitoring for the types of risks identified in the organisation's risk assessment. • Allocation of inadequate significance/weight to a scenario may result in a transaction that should be investigated, not scoring sufficient points to be investigated, and therefore failure to prevent and report money laundering. • Incorrect configuration of the look-back period may result in multiple occurances of alert events not being identified, allocated sufficient weight, and therefore not being investigated.	• Risk assessment should be performed on a periodic basis to identify the risks faced by the organisation. The results of the risk asssessment should be reconciled with the TM system to confirm coverage of key risks. • Risks identified by the risk assessment, proposed scenarios, and thresholds to apply (coverage analysis) should be subject to review, approval, and sign-off by the business. • Adequate documents should be maintained for the approach to segmentation, including data lineage, assumptions, and where segmentation should be reperformed (excessive polarization, orphaned clients, results from below the line testing, etc.). • MI reports should be produced to report client segmentation, changes in segment composition, concentration/polarization, dynamic risk distribution, event alerts distribution by segment, alert aging, alert grouping, identifying clients responsible for multiple alerts, etc. • Client segment, scenarios, thresholds applied, and the relative "weight" or score of an alert should be

	• Incorrect setting of the alert score threshold or auto-suppression parameter (where used) may result in alerts that should be investigated not being subject to investigation. • Inappropriate use of trusted party may result in a transaction that should be investigated, not being allocated sufficient points to be investigated both for single and repeat transactions.	subject to periodic review by business risk for approval. • TM system performance and operations should be subject to review to ensure that scenarios and thresholds are appropriate. Review should include investigation of outlier transactions, transaction distribution, and standard deviation above and below the line testing (ATL/BTL).

Figure 5: Monitoring Risk and Control

Scenario Selection

Red Flags

A red flag is a warning signal that should bring attention to a potentially suspicious situation, transaction, or activity[5] (ACAMS, n.d.). Suspicious situations, transactions, or activity can generate a single red flag or multiple red flags at various stages of the transaction. Recognising a red flag relies, in part, on the adequacy of TM system configuration and the training of staff, but most importantly, appropriate identification of the risks the organisation faces determined by the risk assessment process.

Typologies

The term "typologies" is used to describe the techniques that the criminal is anticipated to use to launder money or finance terrorism. These techniques generate the red flags, which the TM system searches to find unusual or suspicious activity.

Scenarios

The terms "typology" or "scenario" are often used interchangeably, but if we consider that typology describes the technique used by the criminal, then the term "scenario" is the set of rules and threshold values that we use to monitor the transaction for the red flags and typologies.

A scenario therefore identifies the attributes that we want to check and associated threshold values. If the transaction being screened against the scenario is not within tolerance of the threshold value (exceeds or is less than the threshold), it will be allocated a score that reflects the materiality of the event.

For example; a simple scenario rule might be:

> *If total value of a client's inbound wire transfers > 2 x (average total value inbound wire transfers), then score 10 points*

This rule will allocate 10 points when a client receives more than twice the average total value of inbound wires for clients in the same segment. We can decide to allocate additional points if the inbound funds represent greater risk, such as if the wires are from a higher risk geography or if the funds are immediately paid away via an outbound wire and further points if paid away to a higher risk geography, etc. In practice, scenario rules typically evaluate client type, transaction type, transaction value, and timeframe.

[5] ACAMS Glossary of terms - https://www.acams.org/aml-glossary/index-r/#red-flag

The example above is in pseudo-rule format, but the coding logic is similar and should be carefully reviewed to ensure that it reconciles with the system documentation, scenarios, and thresholds.

Scenarios must be maintained to ensure that they remain relevant and address risk. Changes such as how client behaviour evolves and changes in delivery channel may necessitate new scenarios or changes in thresholds. Examples of the change in delivery channel behaviour would be the evolution to greater use of contactless payments and less use of cash, or the use of payment methods such Apple Pay, which may correspond with a reduction in use of cash and debit card transactions.

TM systems are supplied with pre-existing libraries of scenarios, which are typical for the generic financial institution. The implementation process entails selecting which scenarios should be active to monitor for the risks identified by the firm's risk assessment and associated typologies.

A typical method to maintain an understanding of which scenarios should be applied to a given segment is to create a matrix and show the intersections. In the example below, we can see that scenarios 1 is applied to all segments, and there is a difference between Retail Premier and Private banking.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	*Scenario n*
Segment 1 - Retail Basic banking	Y	Y	Y				
Segment 2 - Retail standard banking	Y	Y		Y	Y	Y	
Segment 3 - Retail HNW Premier	Y		Y	Y	Y	Y	Y
Segment 4 - Retail Private	Y			Y	Y	Y	Y

Figure 6: Segment x Scenario Matrix

Selection of scenarios is not a one-off process; scenarios and associated thresholds should be periodically reviewed. Events such as production of a risk assessment, updates to the National Risk Assessment—EU member states are obliged to consider the supranational Risk Assessment and to produce a National Risk Assessment (EU Parliament and Council, 2015)[6]—adverse results from the compliance monitoring process, or typology updates (FATF, n.d.) published by FATF[7] should trigger a review of how the scenarios address the risk.

Scenarios are typically grouped so that they address related groups of typologies. For example, it is typical to have a group of scenarios that evaluate changes in client behaviours, another group that evaluates money laundering using debit cards and cash machines, etc.

[6] DIRECTIVE (EU) 2015/849 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 May 2015 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, and repealing Directive 2005/60/EC of the European Parliament.

[7] https://www.fatf-gafi.org/publications/methodsandtrends/?hf=10&b=0&s=desc(fatf_releasedate)

Scenario groups include:

- High-risk geography – scenarios to evaluate transaction activity with a higher risk geographic nexus
- High-risk entities – scenarios that evaluate transactions performed by a higher risk entity
- Hidden relationships – identification of client, account, and transaction relationships that were not disclosed
- Changes in behaviour – detection of changes in behaviour, such as changes in predominant delivery channel, product utilisation, frequency of international wires, changes in average account balance
- Anomalies in behaviour – detection of behaviour anomalous to the client and segment
- ATM, debit, credit plastics – detection of typologies that exploit the features of plastics, ATM, and POS devices
- Terror financing – detection of client activity that is associated with terror financing

There are numerous sources to identify red flags and typologies that should be considered for inclusion. In addition to the library of scenarios provided by the TM system provider, custom scenarios can be developed to address requirements such as local regulation and specific product risks, etc. Custom scenarios must be carefully reviewed to ensure that they codified correctly and appropriately documented; appropriate scoring is applied and subject to the same governance as standard scenarios.

Financial Action Task Force (FATF) typology reports[8] are an important source of information providing analyses of risk and descriptions and examples of how criminals commit crime and the controls that should be deployed.

TM systems typically produce MI reports that report the number of times a scenario has generated an alert.

The auditor should carefully review this report to identify data such as:

- scenarios that are not generating alerts;
- scenarios that are triggered, but the alerting threshold is not met;
- suspended scenarios (a scenario that is configured to be subject to suppression); and
- "reasonableness" check for scenarios against specific client segments.

Removal of redundant scenarios should be subject to governance to ensure due consideration of the risk and the calibration of thresholds. It may be the case that a scenario appears redundant because it is not generating alerts; this could be in consequence of an incorrectly set threshold. Governance arrangements should include consideration of the risk by business risk and authorisation for systems staff to remove a scenario.

[8] https://www.fatf-gafi.org/publications/?hf=10&b=0&r=%2Bf%2Ffatf_documenttype_en%2Freport&r=%2Bf%2Ffatf_topic_en%2Fmethods+and+trends&q=typologies&s=desc(fatf_releasedate)

Thresholds

Threshold setting is not a one-off exercise that takes place once when the software is installed. Management Information in respect of queues, testing, and detection efficiency should be carefully monitored to identify where thresholds should be reviewed.

As described above, TM systems have default libraries of scenarios that the vendor will select and configure during installation process. The scenarios selected should reconcile with the risk assessment. A key action for the auditor is to check that the installation is risk based and not generic for an institution type; there should be a documented reconciliation between the risk assessment, the scenarios used for monitoring, and the thresholds set.

Scenarios have several parameters that determine how they are applied to a transaction. Parameters include data such as a "transaction size" or "maximum monthly balance," "number of cash deposits," etc. By setting a threshold value, we can score a number of points to reflect the significance of the event. If the accumulated score exceeds the alert threshold, an alert is generated.

Thresholds typically have a set range of permissible values, such as minimum, maximum, and default settings. These are essentially the factory default settings and should be carefully examined by the auditor with appropriate challenge and rationale sought for the use of default values.

Setting the threshold values for certain scenarios is very simple, as the values are defined by local law. For example, if we had a scenario to identify large currency transactions and report them, in certain jurisdictions the threshold for what constitutes a large transaction is in law and therefore easy to set. In this case, the scenario is more a simple rule than a scenario.

Scenarios such as identifying repeat occasional transactions, which, in aggregate, would exceed the occasional transaction threshold and therefore require CDD to be performed, are slightly more complicated to implement.

Where threshold setting gets more interesting is where the values are not mandated. Setting the threshold incorrectly will result in one of two things: firstly, money laundering takes place and is undetected, not investigated and not reported; secondly, a large volume of alerts is generated to the point that a backlog amasses resulting in delays in investigating and preventing known and suspected money laundering/terror financing activity.

If we understand the client segment, we could identify normal behaviour, and, more importantly, outlier or anomalous behaviour. The anomalies may be those transactions that exceed a specific threshold, such as originating transactions that are significantly larger than average or more realistically those that exceed a threshold coupled with another event, which, when considered together, merit alert generation. For example, a small business that has a significant increase in maximum monthly balance may generate an alert; however, if this arises because of a tax refund, then it may not be alerted. Taking this further, if the business receives another refund from the tax authority the following month, it may be indicative of VAT (sales tax/carousel fraud) and therefore should be investigated. These

thresholds and conditional handling form part of the scenario logic; setting the correct threshold values determines the significance or the score given to the transaction.

Initial setting of threshold values can be arbitrary using default settings and local expertise or derived from a statistical approach that focuses on behaviours calculated to identify the statistically anomalous.

A simple form of anomalous behaviour is the outlier transaction. If we select a client segment, it follows that the clients are in that segment because they have similar attributes (assuming the segmentation is correct). Clients with similar attributes would be expected to have similar behaviours. The question becomes: how much deviation from the mean behaviour of other clients in the segment is worthy of investigation?

For example, if we want a scenario to generate an alert if a customer appears to pay away large amounts regularly throughout the month, then we need to calculate the arithmetic mean amount paid away over the month by clients in the same segment. From the mean we can produce bands that correlate to the standard deviations away from the mean.

Standard Deviation Approach

Select the segment that we want to review. This provides a population of clients that by virtue of being within the same segment have similar attributes. By analysing the transactions within the segment, we can calculate values, such as mean transaction size.

1. Select the client segment where we want to check threshold calibration
2. Certain transactions will not be relevant for this purpose, so we dispense with actions such as balance enquires and bank fees
3. If we calculate the total amount paid away and divide this by the total number of transactions, we have the mean. Financial transactions within a segment will typically have a normal distribution, meaning that approximately 50 percent of the transactions will be below the mean, and 50 percent will be above the mean.

$$\text{Mean transaction value} = \frac{\text{Total value of money paid away}}{\text{Total number of transactions}}$$

4. If we show the distribution of transaction and volume of transactions, we obtain the bell curve showing the mean value.

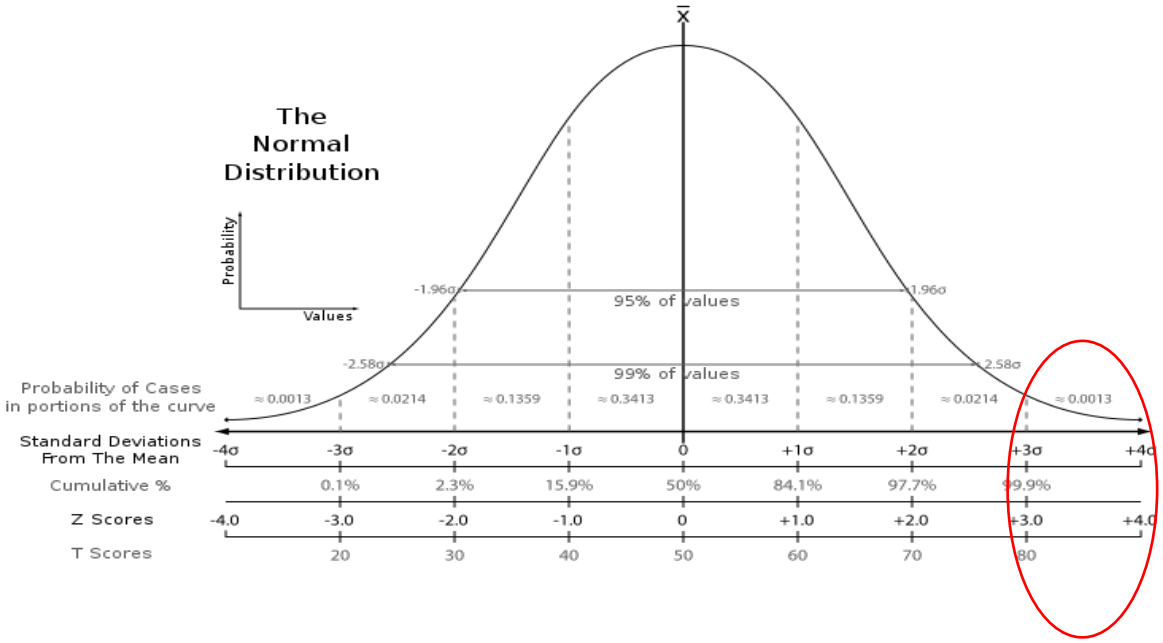


Figure 7: Normal Distribution Curve[9]

The 1st standard deviation from the mean includes 68 percent of all values (34 percent of values less than the mean and 34 percent of values greater than the mean. The 2nd standard deviation from the mean includes 95 percent of all values and the 3rd standard deviation from the mean includes 99.7 percent of all values

5. By calculating the 3rd standard deviation, in addition to seeing the transaction values, which account for 99.7 percent of all transactions, we can see the greatest 0.3 percent of transaction values. The start value of 0.3 percent is a good start point to set the threshold for a scenario to monitor for large transactions.

The TM system documentation may refer to this approach to initial threshold setting as the standard score or as the Z-score. The Z-score is derived by subtracting the mean and then dividing by the standard deviation, thereby producing a standard normal distribution.

$$\text{Z Score} = \frac{\text{Value to be standardised} - \text{Mean value}}{\text{Standard deviation}}$$

[9] https://commons.wikimedia.org/wiki/File:The_Normal_Distribution.svg
Author 'Heds 1 at English Wikipedia/Public domain'

A simple scenario rule using a Z-score, would produce a rule such as:

If client is high risk AND Z-score >= 3 then score 30 points

Clearly the nature of the client base could be that reviewing the 0.3 percent of greatest value transactions is not adequate. It could be that the top 20 percent of accounts have suspicious activity and would only be investigating the top 0.3 percent. We need to perform above and below the line testing to ensure the correct threshold value is used.

Above the Line/Below the Line (ATL/BTL) Testing

We can apply a number of statistical techniques to set the thresholds, but we do not know if the settings are correct unless we perform testing. Testing threshold settings is an iterative approach referred to as above the line (ATL) and below the line (BTL) testing.

Above the line testing is a technique to examine the alerts that scored a given number of points greater than the threshold set to trigger investigation of an alert (consider this the "line"). By manually examining the alerts above the line, i.e., they scored enough points to be investigated, we can check the extent of false alerts. Above the line testing usually commences at 10 percent above the line, so, for example, if alerts scoring 80 points are flagged for investigation, alerts scoring 80 to 88 points are subject to a manual review.

Manual review also permits the auditor to verify the scenario logic and reconcile the points awarded to the different transactions, including exceptional events.

If review concludes that a transaction type is not suspicious, the threshold value can be adjusted to reduce the false positives. Note that above the line testing should be on a transaction-type basis. Is it reliable? One incorrect threshold for a scenario does not mean other thresholds are set incorrectly.

It is important when performing above the line testing that we examine the alert queue MI and consider key metrics, such as:

- ratio of investigated alert to suspicious activity report disclosure;
- alerts generated by scenario; and
- the average score accumulated by scenario, etc.

Above the line testing is especially important where there are backlogs of alerts, as manual review of the constituent factors of the scenario and associated thresholds will ensure optimal detection whilst reducing false positives. This approach to manual review of the transactions above the line should be iterative to set the optimal threshold values and performed periodically when scenarios are reconciled with the risk assessment.

In contrast, below the line testing examines the alerts that have fewer points than the threshold value to trigger an investigation. This is usually performed selecting a percentage, as per above the line testing 10 percent below the investigation threshold value is often used as start position.

Below the line testing is vitally important. It provides the only assurance that we are correctly selecting alerts that must be investigated. If we review alerts that have a score less than the investigation threshold and determine that there is evidence of money laundering/terror financing behaviour, then the threshold is set incorrectly. If we find issues, then it is important that the scenario is examined step by step and the threshold reduced further until review does not identify suspicious transactions that must be investigated.

The auditor should request and examine the results of the below the line testing and pay attention to the iterative process of reviewing the logic and threshold calibration. Where below the line testing does identify that threshold values are incorrect, the auditor should investigate if a look-back was performed, as it is implicit that transactions must have taken place that should have been investigated but did not score sufficient points to be investigated.

The auditor should verify that the rationale for scenario selection, associated thresholds, and threshold testing/calibration decisions are documented and subject to formal change control.

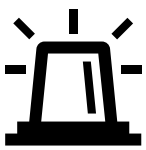
TM systems typically generate log and error files that report scenario and threshold configuration issues. The log usually includes error information, such as:

- Scenarios that do not have screening data available
- Scenarios where thresholds are not set
- Data that is the wrong type or format
- Data values that exceed minimum and maximum thresholds
- Scenarios where the application logic means the scenario will never trigger/generate alert points. Examples of this type of issue include incorrect use of pointers, incorrect nesting of rules, comparing an alphabetic value with a numeric value, etc.

When auditing the TM system, the auditor should ensure that audit log and error report are being produced and that they are subject to review. Because of the operational risk, it is important to carefully consider that the person configuring the system may be the same person reviewing the log of errors.

Alerts

When a transaction is screened and the threshold parameter is exceeded, it is considered an event, and "points" are scored. The number of points should be proportionate to the materiality of the scenario that was triggered. Whereas the transaction has triggered the scenario, this is not yet an alert.

Alerts	Risks	Controls
	• Failure to allocate sufficient point score when a scenario alerts may result in the transaction activity not exceeding the alert investigation threshold, and therefore, potential suspicious activity not being investigated. • Failure to identify suspicious activity or allocate sufficient point score in consequence of focusing on a subset of risk factors • Alerts that should be subject to review are discounted because of incorrect usage of auto-close configurations.	• Active scenarios should be reconciled with the risk asssment to ensure that key risks are being monitored. • Scenarios selected should monitor a range of risk factors, not focus on a limited set of risk factors. • Alert queues should be subject to above the line (ATL) and below the line (BTL) testing to ensure appropriate calibration.

	• Incorrect setting of the scenario look-back period results in transaction patterns and context not being correctly performed, resulting in risk context and mitigants not being appropriately considered. • Service providers fail to adapt procedures in response to changes in risk, resulting in inappropriate closure of alerts.	• Evidence of testing should be maintained with an appropriate degree of independence beween setting threshold values and testing. • Where BTL testing identifies the requirement for updates to scenarios, scenario logic, risk factor focus, threshold values or weight (scenario point score), there must be adequate documentation, evidence of escalation to the anti-money laundering officer, quantification of the issue, and consideration of a look-back.

Figure 8: Alert Events Risk and Control

The threshold to generate an alert is the total point score that must be met (or exceeded) for the event to become an alert.

This should be an important area of focus for the auditor. There is little point in reconciling the risk assessment with the scenarios selected if insufficient weight or points are associated with the event.

In addition to carefully evaluating the appropriateness of the points allocated to events, the auditor should consider the threshold to generate an alert. Documentation should be maintained that provides the rationale for the alert threshold. The rationale must be supported by evidence of testing.

The major risk here is that the alert threshold must not be set to a value that generates a steady workflow to investigators. The alert threshold must be set to ensure that all material events generate an alert. Workflow must be determined by risk, not the available resources. This is a key point that should be examined by the auditor with evidence of the process being examined.

Manipulation of the alert threshold will result in events being ignored or promoted to the alert queue for investigation. Particular attention should be paid where there are initiatives to clear backlogs, outsourcing arrangements, shared service models, or significant changes in the organisations risk profile.

There are, however, several other parameters on a typical TM system that affect the determination to ignore or to generate an alert.

Transaction Look-Back Period

The materiality of an event may be greater if it is a repeated occurrence. For example, if we see an exceptionally large inbound payment, this may trigger an event but not an alert. If the next time we screen transactions, we see another large payment, this would be more material and a greater number of points awarded, potentially resulting in an alert.

The parameter that determines how many days of event history to consider may have a different name on different TM systems, but is the equivalent of a transaction look-back parameter. If the parameter is set to a long period, then there is a greater probability that a previous event will be detected; if the period of time is too short, then the anomalous activity that generated the event may never become an alert and therefore never be investigated. The correct setting will depend upon the nature of the transaction and

behaviour of the client segment. Too great a period will have the effect of normalising the event, and too small will have the effect of not identifying a potential pattern of multiple instances of anomalous activity.

Auto-Close/Alert Suppression

Certain TM systems provide the facility to suppress defined event types. Suppression may be configured for valid reasons, such as the underlying account is the organisation's suspense account, or the parties to the transaction have been subject to EDD measures, and there is a legitimate reason to consider transactions between them to be trusted.

There are various reasons why suppression may be justifiable. The auditor should request details of any suppression rules and ensure that appropriate governance, including rationale, exists for any exceptional treatment of transactions.

Where auto-close/alert suppression is used, the audit review should include evaluating the appropriateness of transaction types selected for suppression. Additional audit checks should include verifying the KYC records for the parties are current, checking the period for which suppression should be applied is appropriate, verifying the authorising senior officer had sufficient information to authorise the 21suppression, confirming the date of next review is set and appropriate, confirming that there is evidence of relationship reviews taking place, etc.

Alert Grouping

Alerts may be linked for a number of reasons. The linking factors can range but are typically because they relate to the same transaction, the same customer, or beneficiary. The TM system may have a feature to perform the grouping so that the point score reflects the materiality of the linked activity, typically resulting in an investigation being performed. A significant benefit of this approach is that the investigator is provided with the linked activity providing more context and therefore enabling a better investigation and quality of disposition.

Where available, alert grouping should be reviewed to ensure optimum alert association, use of alert points, and prioritisation. The points allocated for alerts must reflect the overall materiality.

The logic for use of alert grouping should be reviewed to ensure that where options such as trusted pairs and auto-suppression are used, they do not inappropriately negate the requirement for an investigation. Where a group of alerts includes trusted pair and transactions that would be subject to suppression logic, the suppression logic should be reviewed for appropriateness.

Investigations

Alerts that have sufficient points to meet/exceed the threshold for investigation are flagged to be subject to an investigation. The time the alert is flagged for investigation must be recorded, as there are clear legal and regulatory expectations for the timely investigation, disposition, and, where appropriate, disclosure of suspicious activity.

Investigate	Risks	Controls

	• Failure to recognise activity known to be indicative of money laundering or terror financing behaviour and make an appropriate disclosure (SAR/STR)
	• Failure to correctly identify red flag activity/typology of the money laundering activity and therefore failure to make an appropriate disclosure. This risk manifests in several ways, including:
	  • Failure to appropriately set an account review flag where there is insufficient evidence of money laundering/terror financing to disclose, but sufficient concern that the account activity should be reviewed again in the short term
	  • Failure to perform EDD when an alert is generated, potentially failing to identify context such as connected parties and/or connected beneficial owner/controller, which would result in a disclosure
	• Resources not deployed at areas of greatest risk in consequence of a poorly configured system, scenario selection, threshold setting, and investigtions threshold
	• Failure to investigate and, optionally, submit a Suspicious Activity/Transaction Report (SAR/STR) within the mandated timeframe
		• Investigators must have access to sufficient information to perform a thorough investigation, including:
		  • client file;
		  • account(s) history;
		  • transaction details;
		  • external reference sources;
		  • typology reference sources; and
		  • previous alert history, etc.
		• Investigators must have sufficient ongoing training, experience, and skill to perform the investigation.
		• Investigator performance should be subject to MI reporting to ascertain outlier decision making. Reporting should include:
		  • scenario type investgated;
		  • disposition of alert; and
		  • use of account markers, such as account review flags.
		• Where responsibility for alert investigation has been delegated to another function, business unit, or third party, this must be subject to oversight with appropriate management information reporting and an audit assurance program.
		• Management information should be available including:
		  • alert queue composition by scenario;
		  • alert type and average scores;
		  • alert age profile; and
		  • alert to disclosure ratios.
		• Operating processes should be arranged to enable referrals for internal QA, such as to second-person review.
		• Investigators should follow standard operating procedures for the performance of an investigation ensuring evidence of available data, enquiries made, and decision making.

Figure 9: Investigate Risk and Control

Investigation of alerts requires that a sufficient number of appropriately skilled investigation analysts are provided with data they require to determine if a disclosure must be made to report known or suspected money laundering or terror financing activity.

Legislation typically requires "timely disclosure" as opposed to within specified timeframe; having said this, organisations frequently have backlogs of alerts waiting to be investigated. Prioritising the alert queue to ensure that the most significant issues receive priority is performed by using a combination of factors, including:

- Scenario generating the alert
- Client risk score
- Alert score
- Geography/jurisdiction
- Alert grouping
- The existence of alerts raised manually by staff
- Age of alerts, etc.

Complexity of the alert type may also be used as a factor in prioritisation, as complex cases can expose the institution to significant risk. Investigation of complex cases must only be performed by the most experienced resources. This is to ensure that red flags are appropriately identified, typologies understood, and appropriate tools are used to support an investigation.

Management information is vital to understand the workflow, composition of the queue, proximity of alert scores to the investigation threshold, risk profile, alert aging, disposition outcome, and disclosure ratios. Strategic analysis of the queue can provide insight to trends over time, which should be considered when reviewing which scenarios should be applied and the thresholds that should be set.

When auditing the investigation of TM alerts, the auditor should focus on the capacity of skilled resource to investigate alerts within the mandated timeframe. Workflow data and investigation cases should be examined to ensure that a thorough investigation is being performed, evidence of the investigation is maintained, referral to colleagues takes place and internal records of typologies are updated, and recommendations for trusted pairs, suppression logic, and threshold adjustments are noted.

Conclusion

Transaction Monitoring systems can be complex systems that are left to IT departments to configure and maintain. A "black box" provides a steady stream of transactions that keep a small team of investigation analysts busy.

The auditor's ability to provide independent and objective assurance cannot take place where the auditor does not understand how the black box is configured and how alerts are generated.

The approach described in this paper enables the auditor to develop a conceptual appreciation of the TM system, to be able to ask the right questions, and, importantly, understand the significance of the answers.

When assessing effectiveness, the UK Financial Services Authority (now the Financial Conduct Authority) published a report titled, "Automated Anti-Money Laundering Transaction Monitoring Systems,"[10] in which they state that;

"*to have an effective TM system, we think firms should:*

- *Analyse system performance at a sufficiently detailed level, for example on a rule-by-rule basis, to understand the real underlying drivers of the performance results.*
- *Set systems so they do not generate fewer alerts simply to improve performance statistics. There is a risk of 'artificially' increasing the proportion of alerts that are ultimately reported as suspicious activity reports without generating an improvement in the quality and quantity of the alerts being generated.*

[10] https://www.fca.org.uk/publication/archive/fsa-aml-systems.pdf

- *Deploy analytical tools to identify suspicious activity that is currently not being flagged by existing rules or profile-based monitoring.*
- *Allocate adequate resources to analysing and assessing system performance, in particular to define how success is measured and produce robust objective data to analyse performance against these measures.*
- *Consistently monitor from one period to another, rather than on an intermittent basis, to ensure that performance data is not distorted by, for example, ad hoc decisions to run particular rules at different times.*
- *Measure performance as far as possible against like-for-like comparators, e.g. peers operating in similar markets and using similar profiling and rules."*

The auditor's experience with appreciation of risk and assessment of control adequacy equips them well to apply a conceptual understanding, ask the right questions, and recognise where risks are not being managed as effectively as they should.

Works Cited

ACAMS, n.d. *Glossary of Terms.* [Online]
Available at: https://www.acams.org/aml-glossary/index-r/#red-flag.

EU Parliament and Council, 2015. *DIRECTIVE (EU) 2015/849 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL.* [Online]
Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015L0849&from=EN.

FATF, 2014. *GUIDANCE FOR A RISK-BASED APPROACH THE BANKING SECTOR.* [Online]
Available at: https://www.fatf-gafi.org/documents/riskbasedapproach/documents/risk-based-approach-banking-sector.html?hf=10&b=0&s=desc(fatf_releasedate).

FATF, n.d. *Methods and Trends.* [Online]
Available at: https://www.fatf-gafi.org/publications/methodsandtrends/?hf=10&b=0&s=desc(fatf_releasedate).

FSA, 2007. *Automated Anti-Money Laundering Transaction Monitoring Systems.* [Online]
Available at: https://www.fca.org.uk/publication/archive/fsa-aml-systems.pdf.

Santander Consumer Bank , 2019. *Second quarter report 2019.* [Online]
Available at: https://www.santanderconsumer.no/globalassets/om-oss/investor-relations/financial-reports/santander-consumer-bank-nordic-group-_q2-2019-report.pdf.

Department of Financial Services Superintendent's Regulations Part 504 BANKING DIVISION TRANSACTION MONITORING AND FILTERING PROGRAM REQUIREMENTS AND CERTIFICATIONS
Available at : https://www.dfs.ny.gov/docs/legal/regulations/adoptions/dfsp504t.pdf.

Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency SUPERVISORY GUIDANCE ON MODEL RISK MANAGEMENT OCC 2011-12
Available at : https://www.occ.treas.gov/news-issuances/bulletins/2011/bulletin-2011-12a.pdf.

UK Financial Conduct Authority(FCA) Financial Crime Guide: A firm's guide to countering financial crime risks (FCG)
Available at : https://www.handbook.fca.org.uk/handbook/FCG/2/?view=chapter.

Post-event transaction monitoring process for banks Guidance De Nederlandsche Bank
Available at : https://www.toezicht.dnb.nl/en/binaries/51-236846.pdf.