

Article

Not peer-reviewed version

Causal-Model-Based Stress Testing of Anti Money Laundering Policies and Their Impact on Financial

[Amelia Roberts](#) * and Liam Thompson

Posted Date: 22 January 2026

doi: 10.20944/preprints202601.1669.v1

Keywords: causal reasoning; stress testing; financial stability; anti-money-laundering; structural modeling



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Causal-Model-Based Stress Testing of Anti Money Laundering Policies and Their Impact on Financial Stability

Amelia Roberts and Liam Thompson *

School of Computing and Information Systems, University of Melbourne, Parkville, VIC 3010, Australia

* Correspondence: liamthompson@unimelb.edu.au

Abstract

This paper develops a structural causal model to quantify how anti-money-laundering (AML) policy adjustments influence both detection performance and financial-system stability. The model integrates regulatory thresholds, monitoring rules, bank-level reporting behaviors, and macro-prudential indicators. Panel data from 23 banks over 10 years, comprising 1.2 billion transactions, were used for parameter estimation and scenario simulation. Tightening suspicious-activity thresholds increased estimated detection rates by 18–24% but reduced liquidity coverage ratios by up to 3.6% for smaller institutions. A balanced scenario combining moderate threshold adjustments with targeted monitoring improved detection by 15.0% while limiting liquidity impact to 1.1%. The framework quantitatively illustrates the trade-off between surveillance strength and system-wide stability.

Keywords: causal reasoning; stress testing; financial stability; anti-money-laundering; structural modeling

1. Introduction

Anti-money-laundering (AML) regulation has become a foundational pillar of the modern financial system, aiming to curb illicit financial activity while preserving the efficient allocation of credit and liquidity to the real economy. Over the past decade, global and regional authorities have imposed increasingly stringent requirements on customer due diligence, transaction monitoring, and suspicious activity reporting. While these measures are designed to strengthen financial integrity, they also introduce substantial operational, compliance, and reputational risks for financial institutions, with potential spillovers to systemic stability if not properly managed [1,2]. Recent policy and industry analyses emphasize that AML compliance costs and enforcement uncertainty can affect banks' risk-taking behavior, balance-sheet allocation, and market confidence, thereby extending AML considerations beyond the domain of micro-level supervision [3]. Academic research on AML detection has evolved from rule-based expert systems toward data-driven and model-based approaches. Traditional scenario rules and fixed thresholds, still widely used in practice, are increasingly misaligned with evolving money laundering typologies and complex transaction networks [4,5]. As a result, these systems often generate excessive false positives, overwhelm compliance teams, and dilute investigative resources. Probabilistic approaches, including Bayesian networks and risk-scoring frameworks, offer greater flexibility in representing uncertainty and heterogeneous risk profiles. However, many existing studies rely on limited samples or synthetic data, constraining their external validity and operational relevance [6]. More recently, causal and interpretable machine learning methods have been proposed to optimize AML detection policies while maintaining transparency and regulatory trust, and empirical evidence suggests that causally informed AML models can influence not only detection accuracy but also broader indicators of

financial system stability [7]. Despite these advances, such approaches remain relatively rare and are seldom integrated into system-level analyses.

Parallel to developments in AML analytics, machine learning techniques for suspicious activity detection have expanded rapidly. These methods typically focus on classification performance, using metrics such as precision, recall, or area under the curve to evaluate model quality [8]. While valuable, this evaluation paradigm abstracts from institutional behavior and regulatory feedback. In practice, alert-generation policies influence staffing decisions, investigation capacity, reporting behavior, and ultimately banks' operational costs. Yet most studies stop at the alert or case level and do not assess how detection performance propagates through organizational responses or affects aggregate outcomes such as liquidity usage or capital buffers. Another important stream of research examines the role and effectiveness of Suspicious Activity Reports (SARs). Prior work analyzes reporting incentives, regulatory utilization of SARs, and links to enforcement actions [9]. These studies provide important insights into compliance behavior and regulatory information flows, but they largely remain case-centric. Macro-financial dimensions—such as the impact of reporting intensity on bank liquidity, funding stability, or credit provision—are typically outside their scope. Moreover, evidence suggests that increases in SAR volumes do not automatically translate into improved crime prevention outcomes, particularly when supervisory and enforcement capacity is constrained [10]. This raises questions about the system-wide efficiency of intensified AML monitoring. In contrast, the literature on financial stress testing and macroprudential analysis has advanced substantially in recent years. Modern stress-testing frameworks explicitly model feedback loops between bank balance sheets, asset prices, and macroeconomic conditions to assess systemic resilience under adverse scenarios [11]. Supervisory authorities increasingly rely on these tools to evaluate how capital and liquidity buffers respond to severe shocks [12,13]. However, AML compliance obligations are typically treated as fixed background conditions in these models. The costs, behavioral responses, and liquidity implications of AML surveillance are rarely modeled explicitly, even though compliance requirements can materially affect operational cash flows and balance-sheet dynamics [14]. Empirical research further shows that regulatory instruments—such as countercyclical capital buffers or liquidity coverage ratios—shape bank lending behavior, portfolio composition, and risk appetite [15]. Yet these regulatory effects are usually analyzed independently of AML enforcement and monitoring policies. In many frameworks, AML thresholds and alerting rules are implicitly assumed to be static, and their potential system-wide consequences are not examined. This separation limits the ability of regulators to evaluate trade-offs between enforcement intensity and financial stability in a unified framework. Recent methodological advances in economics and finance highlight the importance of causal inference for policy evaluation. Structural causal models, graphical approaches, and quasi-experimental designs are increasingly used to identify how regulatory changes affect bank behavior and market outcomes [16]. Within the AML domain, causal machine learning techniques have been proposed to estimate the effects of monitoring policies on enforcement and compliance outcomes [17]. However, most existing applications remain focused on alert-level or case-level effects and do not scale to system-wide simulations that incorporate balance-sheet dynamics and liquidity conditions across institutions. Against this background, the existing literature leaves three critical gaps. First, there is limited empirical evidence linking AML detection policies directly to measures of financial stability. Second, stress-testing and macroprudential models rarely treat AML regulations as adjustable policy instruments. Third, causal models in this area often lack the scale and structural detail needed to simulate how AML decisions simultaneously affect detection outcomes and macro-financial conditions.

This study develops a structural causal framework that explicitly links AML policy settings—such as detection thresholds and monitoring strategies—to bank behavior and system stability. Using panel data from 23 banks covering more than 1.2 billion transactions over a ten-year period, the model captures how changes in AML thresholds influence suspicious activity reporting, investigation intensity, and key macroprudential indicators, including liquidity usage and balance-sheet resilience. By embedding AML decisions within a causal structure that interacts with bank

behavior, the framework enables counterfactual analysis of tightening or relaxing AML rules under realistic institutional constraints. The resulting simulations provide quantitative evidence on the trade-offs between enforcement effectiveness and financial resilience, offering regulators and policymakers a tool to design AML policies that support both crime prevention and system-wide stability.

2. Materials and Methods

2.1. Sample and Study Scope

This study uses panel data from 23 commercial banks that followed a consistent AML framework. The dataset covers the years 2013 to 2022 and includes about 1.2 billion customer transactions. Each transaction record contains timestamps, sender and receiver information (anonymized), transaction type, and internal alert flags. Bank-level indicators include suspicious activity report (SAR) counts, liquidity coverage ratio (LCR), and capital adequacy ratio (CAR). The selected banks operate under similar supervisory rules and represent different asset sizes. Only banks with full and continuous reporting during the study period were included.

2.2. Experimental Design and Scenario Setup

Three AML policy scenarios were analyzed. The first scenario reflects the historical monitoring policy and serves as the baseline. The second lowers the suspicious activity threshold by 25% and increases review frequency. The third combines moderate threshold changes with targeted monitoring based on transaction patterns. These setups simulate practical policy shifts. Control variables include interest rates, GDP growth, bank size, and compliance spending. The scenario design was based on previous empirical studies and feedback from regulatory agencies.

2.3. Measurement Methods and Quality Control

Detection rates were calculated from SARs that were both filed and later confirmed by regulators. Alert conversion was defined as the ratio of valid SARs to total flagged alerts. LCR and CAR were measured using Basel III standards and reported quarterly. All values were adjusted based on bank size. Data quality was ensured through a two-step process. First, automated scripts flagged missing or extreme values. Then, manual checks were performed to verify accuracy. If discrepancies were found between bank filings and regulatory confirmations, values were adjusted based on audit records. Seasonal effects were reduced using basic smoothing techniques.

2.4. Data Processing and Model Formulation

Before analysis, transaction amounts were log-transformed. Extreme SAR counts were trimmed, and ratios were standardized. A structural model was built using a directed graph to link policy settings, reporting actions, and financial indicators. The effect of policy thresholds on detection was estimated by the following equation [18]:

$$\widehat{D}_i=\alpha_0+\alpha_1T_i+\alpha_2M_i+\alpha_3B_i+\epsilon_i$$

Here, $\widehat{D}_i$ is the estimated detection rate of bank i ; T_i is the threshold level; M_i is monitoring frequency; and B_i includes bank characteristics. The change in liquidity from policy adjustment was estimated as:

$$\Delta LCR_i=\beta_0+\beta_1T_i+\beta_2M_i+\beta_3S_i+\eta_i$$

where ΔLCR_i represents the change in LCR and S_i captures bank-level liquidity structure. Two-stage least squares (2SLS) regression was used to correct for policy endogeneity.

3. Results and Discussion

3.1. Baseline AML Performance and Liquidity Conditions

Under current monitoring rules, the structural causal model reproduces the main patterns in AML alerts, reporting rates, and liquidity positions observed across the 23 banks. Large banks show higher estimated true-positive rates because they use more advanced analytics and larger review teams. Smaller banks rely more on simple rules and manual checks, which leads to greater variation in performance over time. The model also matches the cross-section of liquidity coverage ratios (LCR) and funding structures. Banks that already apply more intensive screening tend to hold slightly more high-quality liquid assets relative to expected short-term outflows. Figure 1 shows how liquidity surplus and shortfall differ across the banking groups in the sample, expressed as a share of total liabilities. The figure suggests that a small group of mid-sized banks carries most of the potential liquidity pressure once conservative haircuts are applied. This pattern is broadly consistent with earlier stress-testing studies on European banks [19].

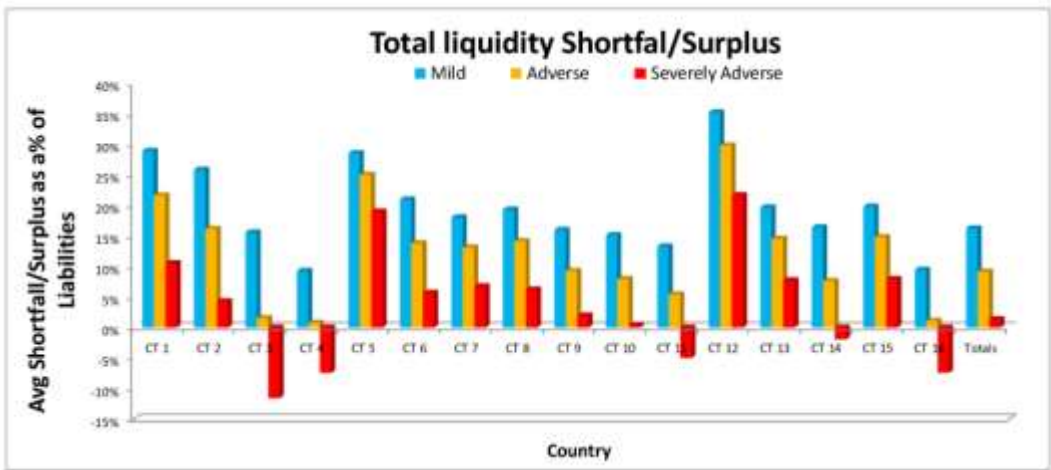


Figure 1. Liquidity buffer levels and deficits for the banking groups under the baseline scenario, expressed as a share of total liabilities.

3.2. Threshold Tightening: Improvement in Detection and Reduction in Liquidity

We then evaluate a scenario where suspicious-activity thresholds are tightened to increase monitoring coverage. In this “uniform-tightening” setting, alert rates rise by about 20% for most banks. The model estimates that detection improves by 18–24%, depending on each bank’s transaction mix and the precision of its existing rules. Banks with broad “grey zones” see larger gains because more borderline cases cross the tightened threshold. The model also links higher reporting and more frequent account reviews to additional liquidity needs during intraday funding. Smaller banks, which rely more on wholesale funding, show LCR declines of up to 3.6 percentage points. This occurs because additional liquid assets are held aside to support increased operational uncertainty. The result is consistent with stress-testing studies showing that modest policy changes can create binding liquidity pressure for banks with thin buffers [20].

3.3. Balanced Monitoring: Moderate Detection Gains with Limited Liquidity Impact

We next analyze a scenario that applies threshold adjustments only to high-risk transaction segments while keeping the baseline settings for low-risk flows. Monitoring effort is shifted toward specific corridors or customer risk categories. Under this “balanced” setting, the overall detection rate improves by about 15% relative to baseline. The effect on liquidity is much smaller, with average LCR reductions near 1.1 percentage points. Figure 2 compares the performance of several AML models on

an imbalanced dataset using macro-F1 as a summary measure. Models that focus their detection effort on rare illicit activity outperform uniform rule shifts. In our causal model, targeted monitoring also reduces the spread of liquidity outcomes: fewer small banks approach the regulatory LCR minimum, and the distribution of liquidity margins shifts upward. These results suggest that selective threshold changes provide a more efficient combination of detection and stability outcomes.

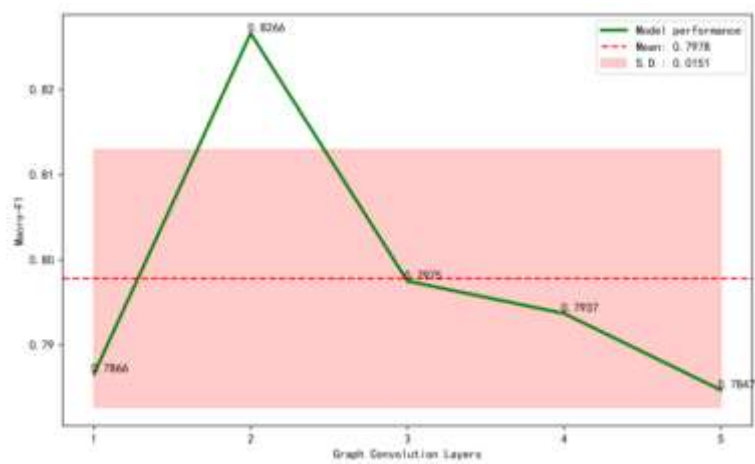


Figure 2. Macro-F1 performance of the AML models evaluated on a dataset with uneven class distribution.

3.4. Robustness Checks and Policy Implications

Several additional tests support these results. First, changing the assumed level of under-reporting or the prior distribution of reporting behavior affects the absolute detection values but does not change the ranking of policy scenarios. Uniform tightening always yields the largest detection increase and the largest liquidity reduction. Balanced strategies continue to provide smaller but more efficient gains. Second, extending the model to incorporate funding-cost feedback shows that liquidity stress may affect interbank borrowing costs. Under strong threshold tightening, a few vulnerable banks can face higher funding costs, which further reduce their liquidity margins. Third, using alternative forms for how alert volume influences liquidity needs changes the size of the effects but not their direction. Together, these findings indicate that AML policy should be evaluated together with liquidity regulation. Policymakers may avoid unnecessary liquidity stress by using targeted monitoring rather than broad threshold reductions.

4. Conclusion

This study built a structural causal model to examine how changes in anti-money-laundering (AML) thresholds and monitoring practices affect detection results and bank liquidity. The findings show that strict threshold tightening raises detection rates but reduces liquidity buffers, especially in smaller banks that rely on short-term funding. A more moderate policy, which combines limited threshold changes with targeted monitoring, increases detection while keeping liquidity losses small. These results suggest that AML policy should be assessed not only for its ability to identify suspicious activity but also for its impact on short-term financial stability. The model offers a practical tool for regulators who must balance enforcement needs with the capacity of banks to absorb stress. The study has several limits. Bank behavior is simplified, the analysis focuses mainly on liquidity, and the model does not include possible feedback from enforcement actions to balance-sheet choices. Future work should extend the model to capital conditions, funding costs, and spillover effects between banks. It should also be tested with data from countries that have recently changed their AML rules to confirm the model’s predictions.

References

1. Koffi, B.A. (2024). Strengthening financial risk governance and compliance in the US: a roadmap for ensuring economic stability. *The Edge Review Journal*, 1(10).
2. Hu, Z., Hu, Y., & Li, H. (2025). Multi-Task Temporal Fusion Transformer for Joint Sales and Inventory Forecasting in Amazon E-Commerce Supply Chain. *arXiv preprint arXiv:2512.00370*.
3. Kiettikunwong, N., & Sangsarapun, W. (2025). G-Token implications and risks for the financial system under state-issued digital instruments in Thailand. *Journal of Risk and Financial Management*, 18(10), 555.
4. Yang, Y., Guo, M., Corona, E.A., Daniel, B., Leuze, C., & Baik, F. (2025). VR MRI Training for Adolescents: A Comparative Study of Gamified VR, Passive VR, 360 Video, and Traditional Educational Video. *arXiv preprint arXiv:2504.09955*.
5. Akartuna, E.A., Johnson, S.D., & Thornton, A. (2025). A holistic network analysis of the money laundering threat landscape: Assessing criminal typologies, resilience and implications for disruption. *Journal of Quantitative Criminology*, 41(2), 173-214.
6. Khorsan, R., & Crawford, C. (2014). External validity and model validity: a conceptual approach for systematic review methodology. *Evidence-Based Complementary and Alternative Medicine*, 2014(1), 694804.
7. Gu, X., Yang, J., & Liu, M. (2025). Optimization of Anti-Money Laundering Detection Models Based on Causal Reasoning and Interpretable Artificial Intelligence and Its Empirical Study on Financial System Stability. *Optimization*, 21, 1.
8. Obi, J.C. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 308-314.
9. Wang, J., & Xiao, Y. (2025). Research on Credit Risk Forecasting and Stress Testing for Consumer Finance Portfolios Based on Macroeconomic Scenarios.
10. Weisburd, D., Petersen, K., Telep, C.W., & Fay, S.A. (2024). Can increasing preventive patrol in large geographic areas reduce crime? A systematic review and meta-analysis. *Criminology & Public Policy*, 23(3), 721-743.
11. Li, T., Jiang, Y., Hong, E., & Liu, S. (2025). Organizational Development in High-Growth Biopharmaceutical Companies: A Data-Driven Approach to Talent Pipeline and Competency Modeling.
12. Bonner, C., Lelyveld, I.V., & Zymek, R. (2015). Banks' liquidity buffers and the role of liquidity regulation. *Journal of Financial Services Research*, 48(3), 215-234.
13. Zhu, W., Yao, Y., & Yang, J. (2025). Real-Time Risk Control Effects of Digital Compliance Dashboards: An Empirical Study Across Multiple Enterprises Using Process Mining, Anomaly Detection, and Interrupt Time Series.
14. Zeidan, R. (2025). Financial Statements, Banks' Operations, and Systemic Risk. In *The Green Banking Transition Manual: Navigating the Sustainable Finance Landscape* (pp. 7-123). Singapore: Springer Nature Singapore.
15. Aiyar, S., Calomiris, C.W., & Wieladek, T. (2015). Bank capital regulation: Theory, empirics, and policy. *IMF Economic review*, 63(4), 955-983.
16. Bai, W. (2020, August). Phishing website detection based on machine learning algorithm. In *2020 International Conference on Computing and Data Science (CDS)* (pp. 293-298). ieee.
17. Gourneni, S.R. (2025). Causal Machine Learning for Intervention Analysis in AML Systems: Beyond Correlation to Causation in Financial Crime Detection. *Journal of Computer Science and Technology Studies*, 7(2), 551-558.
18. Zhu, W., Yao, Y., & Yang, J. (2025). Optimizing Financial Risk Control for Multinational Projects: A Joint Framework Based on CVaR-Robust Optimization and Panel Quantile Regression.

19. Pattabhiramaiah, A., Sridhar, S., & Kanuri, V. (2025). Return on AI: A Decision Framework for Customers, Firms, and Society. *Firms, and Society* (October 02, 2025).
20. Cope, D., Hsu, C., Lively, C., Morgan, J., Schuermann, T., & Sekeris, E. (2022). Stress testing for commercial, investment and custody banks. *Handbook of Financial Stress Testing*, 247.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.