

Beyond predictive accuracy: Sensitivity analysis for behavioural credit assessment



Ioana Florina Coita | Bratislava University of Economics and Business, University of Oradea

Jana Peliova | Bratislava University of Economics and Business

Keywords: Behavioural credit scoring, consumer credit, sensitivity analysis, model governance, AI regulation, open banking

JEL codes: G21, G28, G51, C91, D91

Abstract

Behavioural and AI-based credit assessment increasingly uses transaction histories, repayment sequences and other dynamic signals. Validation usually focuses on predictive accuracy, calibration and fairness, but behavioural variables raise an additional issue: their economic meaning can change with feature definitions, financial state, portfolio scope and disclosure to consumers. This Policy Brief proposes a sensitivity-analysis framework for model governance, drawing on methodological work developed within the WiseCredit research programme while reserving unpublished empirical results for academic papers. The framework emphasises portfolio-level validation, alternative variable constructions, stress and recovery tests, disclosure effects, subgroup stability and cautious treatment of psychometric data.

Disclaimer: This Policy Brief presents a governance framework developed within the WiseCredit project. It intentionally does not report numerical estimates, subgroup associations, repayment-profile results or other novel empirical findings from manuscripts currently under academic review. Acknowledgement. This work was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under project No. 09I03-03-V04-00502.

Five policy messages

- Validate behavioural indicators against the borrower's financial position at portfolio level, not only against behaviour on one account.
- Treat sensitivity analysis as part of model governance: vary feature definitions, thresholds, time windows, coding rules and model specifications.
- If a score is disclosed to consumers, test the score as an intervention because disclosure may alter the behaviour being measured.
- Distinguish financial state, behavioural strategy and personal characteristics; they should not be collapsed into a single risk label.
- Require human review and documented escalation when behavioural indicators and broader financial outcomes point in different directions.

Why behavioural credit variables need a second validation layer

Credit assessment increasingly combines conventional financial information with transaction histories, open-banking data and repeated repayment behaviour. These data can add detail to credit analysis, but they also create a measurement problem. A behavioural variable is rarely a neutral description of one stable object: its meaning depends on how it is defined, the financial state in which it is observed, the time window used and the other credit products held by the borrower.

A point-in-time overdraft balance illustrates the issue. The same balance can reflect a short response to an income shock, persistent dependence on revolving credit, or liquidity used to service another liability. A model that observes only the amount may treat these states as equivalent even though their economic interpretation differs. This is why transparency should start with variable construction rather than only with an explanation generated after a prediction.

There is a second issue when the metric is shown to the person it evaluates. Evidence from consumer credit shows that score information and repayment cues can change subsequent behaviour (Homonoff et al., 2021; Guttman-Kenney et al., 2025). A consumer-facing score can therefore become part of the decision environment. Validation must ask not only whether the measure predicts an outcome, but also whether responding to the measure remains aligned with the financial objective it is intended to represent.

What sensitivity analysis should test

Sensitivity analysis is often treated as a technical robustness exercise performed after a model has been selected. For behavioural credit assessment it should be broader. The object being tested is not only the estimator, but the economic interpretation of the data and the decisions that may follow from the score.

Sensitivity dimension	What to vary	Question for validation	Governance response
Feature construction	Thresholds, ratios, transformations, event definitions and aggregation rules	Does the variable retain the same economic meaning under reasonable constructions?	Document alternatives and investigate material changes in sign, rank or interpretation.
Time and trajectory	Look-back window, persistence rules, shock periods and recovery periods	Is a temporary difficulty distinguished from persistent deterioration or recovery?	Prefer dynamic measures when point-in-time values conceal different paths.
Portfolio scope	Single account versus total debt, liquidity and use of other credit facilities	Can an apparently favourable action on one account be financed by deterioration elsewhere?	Validate against portfolio-level outcomes and cross-product exposure.
Disclosure	Hidden versus consumer-facing score or feedback	Does displaying the metric change the behaviour that generates it?	Treat disclosure as an intervention and monitor adaptation.
Model specification	Feature set, functional form, regularisation, missing-data rule and calibration method	Does the conclusion depend on one modelling choice?	Require stability across plausible specifications or narrow the use case.
Stress and recovery	Income shocks, expense shocks, temporary liquidity constraints and normalisation	Does the signal behave sensibly under stress and after conditions improve?	Test state-dependent error and recovery, not only average performance.
Subgroups and fairness	Relevant protected or vulnerable groups, subject to lawful testing	Are error, calibration or instability concentrated in particular groups?	Escalate material disparities and reassess feature necessity and proportionality.

Feature definitions are policy choices as well as technical choices

How a behavioural variable is coded determines what the model can learn. “Missed payment”, “repayment below obligation”, “overdraft use” and “payment financed by another credit facility” are not interchangeable concepts. They can have different implications for affordability, liquidity and future cost. Data dictionaries should therefore record the economic rationale for each variable, not only its database formula.

The same principle applies to thresholds. A rule based on one month of overdraft use can identify a different behaviour from a rule based on persistence across several months. Sensitivity testing should show whether a risk ranking or policy conclusion survives reasonable changes to these cut-offs. If it does not, the variable should not be treated as a stable behavioural marker.

Trajectories can be more informative than snapshots

Repeated financial data make it possible to distinguish a temporary shock from persistent deterioration. A borrower who uses short-term liquidity and then returns to a stable position is economically different from one whose reliance on high-cost credit continues after the original shock has passed. Model validation should therefore examine persistence, recovery and the sequence of decisions where those features are relevant to the legitimate credit objective.

This also matters for explanation. A statement that risk increased because “overdraft use is high” is less informative than an explanation that identifies persistent overdraft dependence after income or liquidity recovered. The second explanation ties the signal to a financial trajectory that can be checked and challenged.

Portfolio-level validation prevents account-by-account misinterpretation

Households commonly hold several forms of debt at the same time. Research on debt allocation shows that borrowers do not always direct payments toward the option that minimises financing cost (Amar et al., 2011; Gathergood et al., 2019). A behavioural model that rewards repayment intensity on one account should therefore be checked against total debt, debt composition, liquidity and the cost of other borrowing. Otherwise, debt reallocation can be mistaken for improvement.

This does not mean that every model requires complete information on every product. It means that the scope of the claim should match the scope of the data. An account-level model can validly describe behaviour on that account, but it should not automatically be interpreted as a measure of the borrower's overall financial position.

Consumer-facing scores should be validated as interventions

A score that remains internal to a lender is primarily a measurement and prediction tool. A score that is shown to a consumer has another function: it provides feedback and may create a reference point. This distinction matters because a metric can become more salient than the underlying objective. Published research shows that disclosure and repayment prompts can change consumer behaviour, sometimes without a comparable movement in aggregate debt (Homonoff et al., 2021; Guttman-Kenney et al., 2025).

Before a behavioural score is deployed as consumer feedback, validation should identify which actions increase the score and whether those actions improve the legitimate financial target. If the metric rewards regularity, for example, the validation process should not assume that regularity is equivalent to lower indebtedness or lower financing cost. The relationship should be tested directly.

This is also a monitoring issue. Consumer behaviour can change after deployment as people learn how a score responds. Post-deployment validation should therefore track whether the relationship between the score and the intended outcome changes over time, whether new forms of substitution appear and whether the metric remains informative after consumers adapt to it.

Psychometric data require a particularly high threshold

Psychometric information is sometimes proposed as an additional input where conventional credit history is limited. The fact that a psychological measure is correlated with a financial behaviour does not make it an appropriate credit variable. Association, prediction and lawful operational use are separate questions.

A defensible operational use would require evidence that the measure adds out-of-sample value beyond financial and transaction data, remains stable across relevant populations, does not create unacceptable fairness effects, has a clear connection to the legitimate credit objective and can be collected and processed on a sound legal basis. Where these conditions are not met, psychometric measures are better used for research on behavioural mechanisms than as direct penalties or rewards in credit decisions.

This distinction also improves explainability. A consumer can reasonably contest an assessment based on persistent arrears, liquidity stress or use of high-cost credit because these are financial states or behaviours with a direct economic meaning. An explanation based on an inferred personality label is harder to verify, harder to act upon and more likely to blur the boundary between observed financial conduct and personal propensity.

A governance protocol for lenders and model validators

1. **Define the target before selecting behavioural features.** State whether the model is intended to estimate default, arrears, affordability, liquidity stress, financing cost or another legitimate outcome. A score should not be validated primarily against itself.
2. **Create an economic data dictionary.** For each behavioural variable, document the financial state or action represented, alternative constructions, known limitations and the channels through which it could relate to the target.
3. **Run structured sensitivity tests.** Vary thresholds, time windows, transformations, missing-data rules, feature sets and plausible model specifications. Record whether interpretation, calibration, ranking and subgroup performance remain stable.
4. **Test portfolio consistency.** Check whether favourable account-level signals coexist with deterioration in total debt, high-cost borrowing or liquidity. If broader data are unavailable, limit the claim made by the model.
5. **Evaluate disclosure separately.** If the score is shown to consumers, test whether the behaviour encouraged by the score remains aligned with the financial target and whether adaptation changes that relationship.
6. **Stress the model and examine recovery.** Assess behaviour under temporary income or expense shocks and after conditions normalise. A model should distinguish short-lived stress from persistent difficulty when that distinction is relevant.
7. **Define human-review triggers.** Escalate cases in which a behavioural score and broader financial indicators conflict, where model uncertainty is high or where sensitivity tests reveal unstable classification.
8. **Monitor after deployment.** Track data drift, consumer adaptation, cross-product debt transfer, subgroup error and changes in score-outcome alignment. Define thresholds for recalibration, narrowing the use case or withdrawing the model.

Relevance for the EU regulatory framework

These recommendations are consistent with the direction of current EU rules, although the WiseCredit research programme is not itself a legal compliance test. Under the AI Act, AI systems used to evaluate the creditworthiness of natural persons or establish their credit score are listed in Annex III as high-risk, apart from systems used for detecting financial fraud. The high-risk framework addresses risk management, data governance, technical documentation, transparency, human oversight, accuracy, robustness and cybersecurity. Regulation (EU) 2026/1744 moved the application of the relevant Chapter III Sections 1–3 requirements for Annex III high-risk systems to 2 December 2027 (European Parliament & Council of the European Union, 2024, 2026).

The Consumer Credit Directive strengthens procedural rights where creditworthiness assessment uses automated processing of personal data. Consumers can request human intervention, receive a clear and comprehensible explanation, express their point of view and request review of the assessment and credit decision (European Parliament & Council of the European Union, 2023).

The EBA Guidelines on loan origination and monitoring require robust and prudent credit-granting standards and governance around creditworthiness assessment. Sensitivity analysis adds a practical question to these controls: does the economic meaning of a behavioural signal remain stable when reasonable definitions, borrower states and model choices change? For consumer-facing scores, governance should also consider whether disclosure itself alters the behaviour being assessed (European Banking Authority, 2020).

Ongoing WiseCredit research

WiseCredit develops controlled repeated credit-repayment experiments to study how borrowers respond to common financial conditions, shocks, liquidity constraints and different forms of feedback. The programme also examines how behavioural, transaction and psychometric information should be defined and validated before it is considered for credit assessment.

The present Policy Note is deliberately methodological. Detailed estimates from the WiseCredit experiments, including treatment effects, repayment-profile analyses, psychometric associations and robustness results, are reserved for separate academic manuscripts currently under review. This separation allows the policy discussion to focus on a transferable governance framework without pre-empting the original empirical contribution of those papers.

Conclusion

Behavioural data can make credit assessment more informative, but they can also make validation more difficult. This type of variable may change meaning with its threshold, time window, financial context or portfolio scope. When a score is shown to consumers, the metric can also influence the conduct it is intended to measure.

For lenders, model validators and supervisors, the practical implication is to move sensitivity analysis from a technical appendix into the model-governance process. The key tests are whether the economic interpretation survives reasonable alternative definitions, whether account-level signals remain consistent with the borrower's broader financial position, whether the model behaves sensibly under stress and recovery, and whether disclosure changes the score-outcome relationship.

A transparent credit model should be able to explain not only why a variable predicts risk, but also what the variable represents, how stable that meaning is and what happens when people respond to the metric. Those questions become more important as credit assessment uses richer behavioural data and as AI systems move closer to the consumer decision process.

References

- Amar, M., Ariely, D., Ayal, S., Cryder, C. E., & Rick, S. I. (2011). Winning the battle but losing the war: The psychology of debt management. *Journal of Marketing Research*, 48(SPL), S38–S50. <https://doi.org/10.1509/jmkr.48.SPL.S38>
- Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7), 2845–2897. <https://doi.org/10.1093/rfs/hhz099>
- European Banking Authority. (2020). Guidelines on loan origination and monitoring (EBA/GL/2020/06). <https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/guidelines-loan-origination-and-monitoring>
- European Parliament & Council of the European Union. (2023). Directive (EU) 2023/2225 of 18 October 2023 on credit agreements for consumers and repealing Directive 2008/48/EC. *Official Journal of the European Union*, L 2023/2225, 30.10.2023. <https://eur-lex.europa.eu/eli/dir/2023/2225/oj>
- European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). *Official Journal of the European Union*, L 2024/1689, 12.7.2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- European Parliament & Council of the European Union. (2026). Regulation (EU) 2026/1744 of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the

implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). Official Journal of the European Union, L 2026/1744, 24.7.2026. <https://eur-lex.europa.eu/eli/reg/2026/1744/oj>

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, 77(1), 5–47. <https://doi.org/10.1111/jofi.13090>

Gathergood, J., Mahoney, N., Stewart, N., & Weber, J. (2019). How do individuals repay their debt? The balance-matching heuristic. *American Economic Review*, 109(3), 844–875. <https://doi.org/10.1257/aer.20180288>

Guttman-Kenney, B., Adams, P., Hunt, S., Laibson, D., Stewart, N., & Leary, J. (2025). The semblance of success in nudging consumers to pay down credit card debt. *American Economic Journal: Economic Policy*, 17(4), 72–105. <https://doi.org/10.1257/pol.20230568>

Homonoff, T., O'Brien, R., & Sussman, A. B. (2021). Does knowing your FICO score change financial behavior? Evidence from a field experiment with student loan borrowers. *The Review of Economics and Statistics*, 103(2), 236–250. https://doi.org/10.1162/rest_a_00888

About the author(s)

Ioana Florina Coita, PhD, is a researcher in the Department of Finance at the Bratislava University of Economics and Business and a lecturer in the Department of Finance and Accounting at the University of Oradea. Her research focuses on behavioural finance, AI and machine learning in finance, creditworthiness assessment, explainability and responsible use of financial data. She coordinates the WiseCredit research project on behavioural and open-banking approaches to credit assessment.

Jana Peliova, Assoc. Prof., PhD, is an associate professor in the Department of Finance at the Bratislava University of Economics and Business and Vice-Rector for Management of Academic Projects. Her research focuses on public finance, behavioural economics, experimental methods and behavioural aspects of financial and public-policy decisions. She supervises the research activities inside WiseCredit project.

SUERF Policy Notes and Briefs disseminate SUERF Members' economic research, policy-oriented analyses, and views. They analyze relevant developments, address challenges and propose solutions to current monetary, financial and macroeconomic themes. The style is analytical yet non-technical, facilitating interaction and the exchange of ideas between researchers, policy makers and financial practitioners.

SUERF Policy Notes and Briefs are accessible to the public free of charge at <https://www.suerf.org/publications/suerf-policy-notes-and-briefs/>

The views expressed are those of the authors and not necessarily those of the institutions the authors are affiliated with.

© SUERF – The European Money and Finance Forum. Reproduction or translation for educational and non-commercial purposes is permitted provided that the source is acknowledged.

Editorial Board: Ernest Gnan, David T. Llewellyn, Donato Masciandaro

Designed by the Information Management and Services Division of the Oesterreichische Nationalbank (OeNB)

SUERF Secretariat

c/o OeNB, Otto-Wagner-Platz 3A-1090 Vienna, Austria

Phone: +43 1 40 420 7206 E-mail: suerf@oenb.at Website: <https://www.suerf.org/>

ISSN 2791-5395 (Online)